

Hydrogen bonding in acrylamide and its role in the scattering behavior of acrylamide-based block copolymers

Elena Patyukova, Taylor Rottreau, Robert Evans, and Paul D. Topham
Aston Institute of Materials Research, Aston University, Birmingham, B4 7ET, UK*

Martin J. Greenall

School of Mathematics and Physics, University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, UK[†]

(Dated: May 28, 2018)

Hydrogen bonding plays a role in the microphase separation behavior of many block copolymers, such as those used in lithography¹, where the stronger interactions due to H-bonding can lead to a smaller period for the self-assembled structures, allowing the production of higher resolution templates. However, current statistical thermodynamic models used in descriptions of microphase separation, such as the Flory-Huggins approach, do not take into account some important properties of hydrogen bonding, such as site specificity and cooperativity. In this combined theoretical and experimental study, a step is taken toward the development of a more complete theory of hydrogen bonding in polymers, using polyacrylamide as a model system. We begin by developing a set of association models² to describe hydrogen bonding in amides. Both models with one association constant and two association constants are considered. This theory is used to fit IR spectroscopy data from acrylamide solutions in chloroform, thereby determining the model parameters. These parameters are then employed to calculate the scattering function of the disordered state of a diblock copolymer with one polyacrylamide block and one non-hydrogen-bonding block in the random phase approximation. It is then shown that the expression for the inverse scattering function with hydrogen bonding is the same as that without hydrogen bonding, but with the Flory-Huggins parameter χ replaced by an effective value $\chi_{\text{eff}} = \chi + \delta\chi_{\text{HB}}(f)$, where the hydrogen-bonding contribution $\delta\chi_{\text{HB}}$ depends on the volume fraction f of the hydrogen-bonding block. We find that models with two constants give better predictions of bond energy in the acrylamide dimer and more realistic asymptotic behavior of the association constants and $\delta\chi_{\text{HB}}$ in the limit of high temperatures.

I. INTRODUCTION

Hydrogen bonding interactions occur very widely in nature. Although individual H-bonds are relatively weak, their effect on the physical properties of substances can be profound, and is responsible for the anomalous properties of water and the secondary structure of proteins. However, the characteristics of hydrogen bonding, such as site specificity and cooperativity, make it difficult to build a general theoretical description of H-bonding systems³.

One of the most natural ways to describe the thermodynamics of the formation of hydrogen bonds is to treat this phenomenon as a reversible chemical reaction. This *association model* approach (sometimes called the ERAS model⁴) was initially proposed to describe hydrogen bonding association in alcohols and polyalcohols². In the framework of this model, it is assumed that alcohols in the liquid state form a full range of linear chain aggregates due to hydrogen-bonding association. It was shown that chemical equilibrium in this kind of system is described well by two association constants, one corresponding to the formation of a dimer (dimer association constant) and the other corresponding to addition of further molecules to the chain (multimer association constant).

The association model of hydrogen bond formation in alcohols was later applied by Painter and Coleman to describe the miscibility of hydrogen-bonding

homopolymers^{2,5}. They showed that association constants measured for low molecular weight analogs of polymer segments can be used (after rescaling in order to take into account the difference between the molar volumes of monomers and polymer segments) to describe hydrogen bonding in polymer systems. This approach has several strengths: (i) its parameters are measurable quantities, (ii) it treats non-hydrogen-bonding interactions and hydrogen-bonding interactions separately, (iii) the number of hydrogen bonded contacts is not random, and (iv) it works as an extension of the Flory-Huggins theory of polymer melts, which is the basic theoretical platform in polymer physics⁶. However, as this work was focused specifically on alcohols, hydrogen bonding in homopolymers and diblock copolymers is still often described by means of a negative Flory-Huggins parameter^{7-9,10}.

Taking into account the virtues of the association model approach, it is useful to extend it to other classes of self-associating hydrogen-bonding compounds such as amides and acids. Here, we develop a set of association models for amides and test them by comparison with IR absorption measurements on acrylamide solutions.

The choice of acrylamide as a system to study is motivated by the current interest in the properties of its corresponding polymer, polyacrylamide. Polyacrylamide is a commercially important polymer which, in addition to its uses in chromatographic columns, soft contact lenses and cosmetics, is now finding applications in the areas of biomaterials and smart materials research^{11,12}. A key

factor in these applications is hydrogen bonding: the acrylamide group has both hydrogen donor and acceptor sites and can serve as a universal hydrogen bonding agent.

In addition to the strengths of the association model listed above, we also believe that it will yield insights into hydrogen-bonded acrylamide aggregates that would be difficult, or even impossible, to obtain by other techniques, such as density functional theory (DFT)^{13,14} and molecular dynamics simulations¹⁵. DFT is a powerful tool that gives many valuable insights into the physics of hydrogen bonding, such as the effect of the conformation and relative positions of the molecules on the energy of the hydrogen bonding interaction. It can also be used to investigate hydrogen-bonded clusters, and such studies have been carried out for acetamide^{16,17}. However, to the best of our knowledge, existing studies on acrylamide focus on the structure and spectral features of isolated molecules and hydrogen-bonded dimers and do not provide any information on networks of hydrogen bonds or on the entropy of hydrogen bond formation in an ensemble of molecules.

Molecular dynamics (MD) has also been used to study hydrogen-bonded networks and, for example, has been used to investigate liquid formamide¹⁵. However, there are questions about the use of MD in these systems, since it was shown in the case of alcohols that molecular dynamics simulations using common force fields do not reproduce the spectral features of hydrogen-bonded aggregates in solution, even on a qualitative level¹⁸. Furthermore, there is always a constraint on the size of the system in molecular dynamics simulations, which limits the possibility of simulating the true size distribution of aggregates in solution.

The main difficulty in constructing an association model of acrylamide is that there is not enough information about the “rules” of association and the minimal number of association constants necessary to describe association. In the literature of the interpretation of IR data from solutions of amides, all models that we are aware of use linear chain aggregates or cyclic dimers¹⁹. In the crystal phase, acrylamide is known to form two-dimensional ribbon-like networks of hydrogen bonds²⁰ with two bonds per oxygen, two bonds per NH₂ group and ribbons built from the double-bonding of cyclic dimers. However, we believe that, in both the solution and the melt, the association model may be different. For example, for relatively large acetamide clusters¹⁷ with aggregation numbers up to $i = 15$, it was shown by means of DFT simulations that clusters with “irregular” (as well as linear and cyclic) structure have lower energies than clusters constructed from crystal polymorphs, and we expect similar behavior for acrylamide.

Our strategy to deal with this uncertainty is to develop a set of models with different association rules. We start with models with one association constant (in other words, in all these models we assume that all hydrogen bonds have the same energy regardless of their position

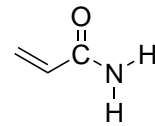


FIG. 1: Acrylamide molecule.

inside the aggregate). These models are then applied to interpret IR spectroscopy data from acrylamide solutions in chloroform and determine the association constants. Next, models with two association constants are investigated. These give better correspondence to the bond energies in acrylamide dimers predicted by DFT^{13,14}.

At the end of the paper, we take a first step toward the application of the association model approach to the description of hydrogen bonding in block copolymers by calculating the structure factor of a disordered melt of block copolymer with one hydrogen-bonding self-associating block and one non-hydrogen-bonding block in the random phase approximation (RPA). Next, the models and parameters determined for acrylamide are used to estimate how the Flory-Huggins parameter that appears in the RPA formula for the structure factor of the diblock copolymer with a polyacrylamide block is shifted by the presence of hydrogen bonds in these systems. It should be noted that these results can be considered only as preliminary, because, though our calculations take into account the non-randomness of hydrogen bonding contacts, they do not take into account the non-randomness of mixing in polymer systems, which should be accounted for in the future.

II. MODELS WITH ONE ASSOCIATION CONSTANT

According to one definition, a hydrogen bond is an attractive short-ranged force between a hydrogen atom bonded to a strongly electronegative atom and another electronegative atom with a lone pair of electrons.

Acrylamide has a primary amide group (see Figure 1), which consists of an oxygen atom (a hydrogen acceptor) and an NH₂ group (a hydrogen donor). The oxygen atom can potentially form two hydrogen bonds as it has two lone electron pairs. Since it has two hydrogens, the NH₂ group can also potentially form two bonds. However, it is also possible that the oxygen atom will predominantly form only one bond (as in alcohols)²¹ or that the second hydrogen on the NH₂ group will lose its donating properties after the first hydrogen becomes bonded²². Different rules of association produce aggregates of different architectures. If we allow only one bond per oxygen and one bond per NH₂ group, then the hydrogen-bonding association of acrylamide molecules produces linear chains. In all other cases, branched aggregates are produced.

From both experimental studies and DFT simulations it is known that tautomerism is not present in acry-

TABLE I: List of models with one association constant and no cycles.

Model	association rules
0	only linear dimers
1	one bond per oxygen, two bonds per NH ₂ group
2	two bonds per oxygen, two bonds per NH ₂ group
3	one bond per oxygen, one bond per NH ₂ group
4	two bonds per oxygen, one bond per NH ₂ group

lamide because imidic acid has an energy approximately 11 kcal/mol higher than the ground state energy of the syn-isomer¹⁴, so we do not consider the possibility of tautomerism in our models. Isomerism is also not considered, because 95% of the molecules are in the syn-isomer state at room temperature due to the large difference in ground state energy between the two isomers²³. In all of our models, it is also assumed that the two hydrogens are equivalent and the slight difference in electro-negativity between them is neglected.

In this section, we consider models for which it is assumed that all hydrogen bonds have the same energy regardless of their position in the aggregate and that there are no cycles of any kind. The list of models generated by these assumptions and different association rules is presented in Table I.

In order to illustrate our calculation method, model 1 is used as an exemplar. The calculations for all other models can be found in the supplementary materials.

In model 1, it is assumed that the rules of association can be formulated as “one bond per oxygen, two bonds per NH₂ group”. The starting assumption is that the free energy of the solution can be written as

$$F = F_0 + F_{\text{HB}}, \quad (1)$$

where F_0 is the free energy of solution without any hydrogen bonding association of the dissolved molecules and F_{HB} is the contribution due to hydrogen bond formation.

If there are N acrylamide molecules in solution with M hydrogen bonds in total between them and the energy of one hydrogen bond is ϵ , one can write²⁴

$$F_{\text{HB}} = M\epsilon - kT \ln [p^M \Xi], \quad (2)$$

where p is the probability of formation of one bond, which can be expected to be inversely proportional to the volume of the system, so that $p = C/V$ (where C is a constant), and Ξ is the combinatorial number of ways to form M bonds in the system. It is worth noting that here we implicitly use our assumption about the absence of cycles because we suppose that the formation of each bond contributes to the statistical weight factor p , which accounts for the loss of entropy due to bond formation. In the case when an additional bond is formed that leads to the creation of a cycle, the entropy loss is either absent or much smaller because the participating molecules are

already held close to each other by other bonds in the aggregate.

The expression for Ξ in the case when one bond per oxygen and two bonds per NH₂ group is allowed has the form

$$\Xi = \frac{N!2^M}{(N-M)!} \frac{(2N-2)!}{(2N-2-M)!} \frac{1}{M!}. \quad (3)$$

Here, the first factor is the number of ways to choose M acceptors for M bonds out of N molecules, taking into account that each oxygen can form only one bond, but has two bonding sites. The second factor is the number of ways to choose M donors for the bonds (at this stage all atoms are treated as distinguishable) and the last factor accounts for the indistinguishability of bonds. Since M and N are large numbers, the -2 terms in the second factor will be neglected.

Substituting Ξ into Equation 2, using Stirling’s formula and minimizing F_{HB} with respect to M yields

$$\frac{M}{2(N-M)(2N-M)} = \frac{K}{V}, \quad (4)$$

where the equilibrium association constant $K \equiv C \exp(-\frac{\epsilon}{kT})$ has been introduced. Alternatively, in terms of concentrations

$$\frac{m}{2(n-m)(2n-m)} = K. \quad (5)$$

Solving this quadratic equation with respect to m , the dependence of the free energy on the total concentration of solution can be obtained as

$$\frac{F_{\text{HB}}}{kTV} = m + n \ln \left[\frac{(n-m)(2n-m)^2}{4n^3} \right], \quad (6)$$

where

$$m = \frac{1 + 6Kn - \sqrt{1 + 12Kn + 4K^2n^2}}{4K}. \quad (7)$$

In order to couple this expression with the Flory-Huggins formula for the free energy of the system without hydrogen bonds, it can be rewritten in terms of volume fractions as

$$\frac{F_{\text{HB}}v}{kTV} = \phi_m + \phi \ln \left[\frac{(\phi - \phi_m)(2\phi - \phi_m)^2}{4\phi^3} \right], \quad (8)$$

with

$$\phi_m = \frac{1 + 6K'\phi - \sqrt{1 + 12K'\phi + 4K'^2\phi^2}}{4K'}, \quad (9)$$

where v is a reference volume and $K' = K/v$ is a dimensionless association constant. As the calculation method is the same for all models, only the final expressions for the free energies of the other models with one association constant are presented here, in Table II. The full derivations can be found in the Supporting Information.

TABLE II: Free energies for models with one association constant.

Model	F_{HBV}/kTV	ϕ_m
0	$2K'\phi_1^2 + \phi \ln \frac{\phi_1}{\phi}$	$\frac{1+16K'\phi - \sqrt{1+32K'\phi}}{32K'}$
1 & 4	$\phi_m + \phi \ln \left[\frac{(\phi - \phi_m)(2\phi - \phi_m)^2}{4\phi^3} \right]$	$\frac{1+6K'\phi - \sqrt{1+12K'\phi + 4K'^2\phi^2}}{4K'}$
2	$\phi_m + 4\phi \ln \left(1 - \frac{\phi_m}{2\phi} \right)$	$\frac{1+4K'\phi - \sqrt{1+8K'\phi}}{2K'}$
3	$\phi_m + 2\phi \ln \left(1 - \frac{\phi_m}{\phi} \right)$	$\frac{1+8K'\phi - \sqrt{1+16K'\phi}}{8K'}$

Model 4 is similar to model 1 but with the roles of donors and acceptors exchanged, and is therefore described by the same equations. However, in this model, the number of free NH_2 groups will be different from model 1, so it is considered as a separate case here.

III. DETERMINATION OF ASSOCIATION CONSTANTS

The association constants of alcohols were previously measured by others using IR spectroscopy with the help of the following idea^{25,26}. Suppose that the hydrogen-bonding substance is dissolved in a solvent that has no specific (i.e. hydrogen bonding or strong polar) interactions with the solute. Then, at vanishingly small concentrations of solute, peaks corresponding to the vibrations of the hydrogen-bonding groups in isolated molecules should be seen. As the concentration is increased, new peaks should appear that correspond to hydrogen-bonded states of hydrogen-bonding groups, as hydrogen bonding changes the absorption frequency of groups participating in the bond. In consequence, the dependence of the height of the peaks corresponding to absorption by isolated molecules should have a weaker than linear dependence on the total concentration of the solution. So, if a formula can be found to describe how the concentration of the species corresponding to a given peak depends on the total concentration, then the association constants can be determined by fitting this expression to experimental results on the dependence of the peak height on the total concentration, with the association constants treated as adjustable parameters.

This procedure is now applied to the case of solutions of acrylamide in chloroform. Chloroform is chosen because it is a non-hydrogen bonding solvent that dissolves acrylamide sufficiently well to give a good range of concentrations (compared, for example, to carbon tetrachloride), and because it has a relatively high boiling temperature (compared, for example, to dichloromethane), to allow the measurements to be conducted over a sufficiently broad range of temperatures.

Figure 2 shows the changes of IR absorption by acrylamide in the range 3700 cm^{-1} to 3000 cm^{-1} as the total concentration of the solution is increased. At low concentrations, we can see two clear peaks at 3414 cm^{-1} and

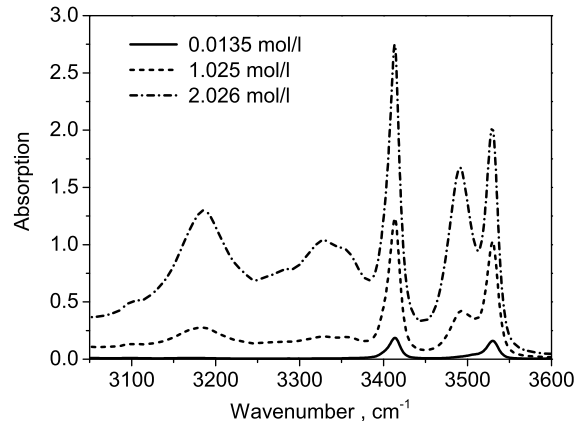


FIG. 2: IR spectrum of acrylamide in chloroform at 22°C at different concentrations focusing on the NH_2 absorption region of spectrum for acrylamide.

3530 cm^{-1} , which are attributed to the in-phase and out-of-phase vibrations of the NH_2 group respectively¹³. It should be noted that, even at the lowest concentrations we studied, there is a shoulder on the 3530 cm^{-1} peak, which we are unable to assign accurately. At larger concentrations, the peaks at 3414 cm^{-1} and 3530 cm^{-1} remain, but the shoulder develops and a broad conglomerate of a number of peaks at lower energies (to the left of 3414 cm^{-1}) appears. It is well known that hydrogen bonding of the donor N-H group leads to a red-shift of the N-H vibration frequency from that of the free group²⁷, so all peaks that appear as the concentration of acrylamide solution increases are assigned to absorption by NH_2 groups in different hydrogen-bonded states.

However, one must be cautious with respect to the attribution of the 3414 cm^{-1} and 3530 cm^{-1} peaks to the “free” NH_2 groups, because the bonding state of the oxygen in the same amide group can influence the frequency of absorption. Extensive attempts to rationalize the rules according to which hydrogen bonding affects the absorption wavelengths in amides resulted in a conclusion that there are no universal rules and that each amide system should be carefully studied in order to find the contributions to the shifts in each particular case²⁸⁻³¹. DFT simulations of hydrogen bonded aggregates of acrylamide could potentially have shed some light on this matter, but we are not aware of any such work in the literature, and the closest study we could find was carried out for perfluorinated polyamides²². In their work, the shifts in absorption wavenumbers of the NH_2 group in linear and cyclic dimers and trimers were calculated. Interestingly, the shift of the out-of-phase vibrations of the free NH_2 groups in the linear dimer was calculated to be 6 cm^{-1} and in the linear trimer to be 15 cm^{-1} . When applied to the spectra of acrylamide, this would result in a contribution of the free groups of linear dimers to the 3530 cm^{-1} band but would also mean that the absorp-

tion of the free group in the trimer would lie away from this band. Similar behavior was reported for acetamide clusters¹⁶. Consequently, based on the available information, the following possibilities for peak attribution have been considered. The first possibility is that the “free” peaks correspond to unimers. This assumption implies that any bonding of oxygen in the amide group substantially shifts the absorption of the NH₂ group. The opposite possibility is that the “free” peaks correspond to the free NH₂ groups regardless of the bonding state of oxygen in the same amide group. Finally, the third possibility considered is that these peaks correspond to the absorption of the free NH₂ groups in only unimers and dimers, which would correspond to the case when the shift of absorption of a free NH₂ group in a dimer is small enough to give a contribution to the 3530 cm⁻¹ peak together with free molecules, but the shift of absorption in free NH₂ groups in larger aggregates is large enough not to give a contribution to the 3530 cm⁻¹ peak.

Returning to model 1, one can find now the dependence of the concentration of free molecules and free groups on the total concentration of the solution.

Let us first find the dependence of the total concentration on the concentration of free molecules. Hydrogen bonding is described in the current work as a reversible chemical reaction that produces a range of aggregates of different structures and sizes, and it is assumed that all aggregates are tree-like and no cycles can be formed. In this case, the concentration of aggregates of size i has the form $c_i = \alpha_i K^{i-1} c_1^i$ where K is the equilibrium association constant, c_1 is the concentration of free molecules and α_i is a coefficient that depends on the size of the aggregate. With knowledge of c_i , the total concentration of the solution n and the concentration of bonds m can be calculated to be

$$n = \sum_{i=1}^{\infty} i c_i = c_1 \sum_{i=1}^{\infty} i \alpha_i (K c_1)^{i-1} \quad (10)$$

and

$$m = \sum_{i=1}^{\infty} (i-1) c_i = c_1 \sum_{i=1}^{\infty} (i-1) \alpha_i (K c_1)^{i-1}. \quad (11)$$

Next, the function

$$g(x) = \sum_{i=1}^{\infty} \alpha_i x^i \quad (12)$$

is introduced, where $x = K c_1$; then, $K m = x g' - g$, $K(n - m) = g$ and $K(2n - m) = x g' + g$. Substituting these expressions into Equation 5 yields

$$x g'(x) - g(x) = 2g(x)(x g'(x) + g(x)). \quad (13)$$

The solution of this equation that remains finite as $x \rightarrow 0$ is³²

$$g(x) = \frac{\exp A - 4x - \sqrt{\exp 2A - 8x \exp A}}{8x}, \quad (14)$$

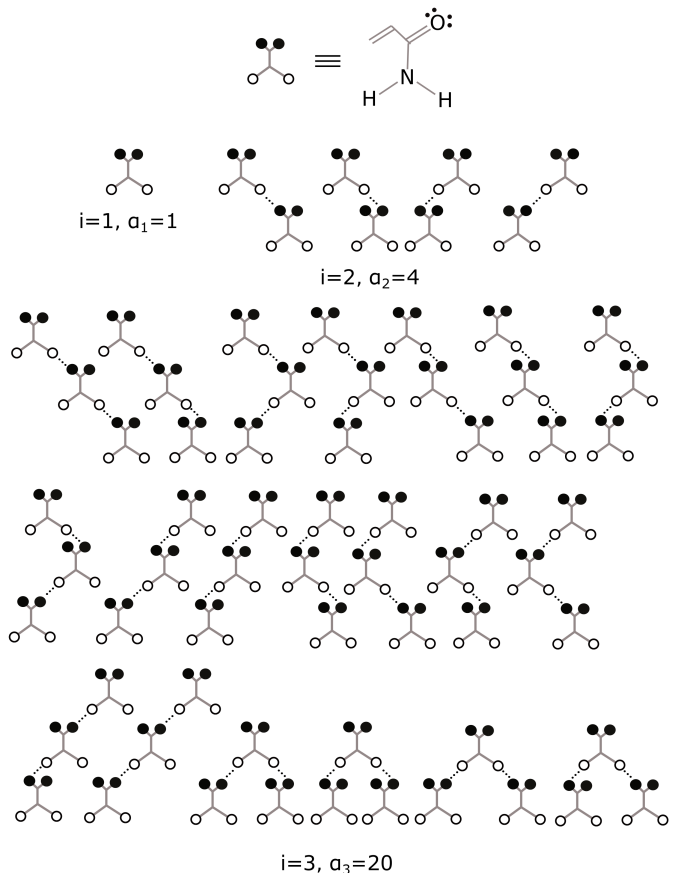


FIG. 3: Schematic representation of aggregates with size up to $i = 3$ in model 1.

where A is a constant determined by the boundary conditions (i.e. the value of α_1 , which is put everywhere equal to 1). We can find α_i by expanding Equation 14 as a Taylor series. The general formula for α_i is

$$\alpha_i = 2^{i-1} \cdot \frac{(2i)!}{i!(i+1)!}. \quad (15)$$

The sequence $\beta_i = \frac{(2i)!}{i!(i+1)!}$ is known as the Catalan numbers (published electronically at <https://oeis.org>, May 2018). It is known that the Catalan numbers represent the number of different rooted binary trees with $i + 1$ leaves. In the current case, there is an additional factor of 2^{i-1} , since each molecule apart from the root can be added in two ways to form a bond with one of the free hydrogens as there are two bonding sites on the oxygen. It can then be said that the physical meaning of α_i is the number of ways to compose an aggregate of size i out of i molecules. All aggregates allowed in our model with size up to $i = 3$ are shown in Figure 3.

Substituting the expression for α_i into Equation 10, a relation between the concentration of free molecules and

the total concentration of the solution is obtained:

$$n = \frac{1 - 4c_1K - \sqrt{1 - 8c_1K}}{8c_1K^2\sqrt{1 - 8c_1K}}. \quad (16)$$

This expression was used to fit our IR spectroscopy data in the case where the 3530 cm^{-1} peak is attributed to free molecules, because in this case we simply write $c_1 = Ax$ where A is some constant and x is a peak height.

It is straightforward to obtain from this expression a fitting equation for the case where the peak is attributed to both free molecules and free groups in dimers. In this case $Ax = c_1 + 4Kc_1^2$.

Finally, the case where the peak is attributed to free NH_2 groups needs to be considered. In order to find the fitting equation, it is necessary to find the concentration of free groups. However, the number of free groups depends on the structure of the aggregate. The most convenient way to do these calculations is to consider the model with two association constants determined by the bonding state of the NH_2 group in the donor molecule (see Figure 4) and, after the calculations are complete, put the association constants equal to each other. So, we will assume that the energy of the bond is ϵ_1 in the case when a second proton in a donor molecule is free and ϵ_2 otherwise. For the sake of clarity, further calculations are omitted here (these can be found in the supporting information in Section 1.6), and the final result for the dependence of the total concentration n on the number of free groups n_f is

$$n = \frac{1 - 4Kn_f + 16K^2n_f^2 - (1 - 4Kn_f)\sqrt{1 + 16K^2n_f^2}}{8K^2n_f}. \quad (17)$$

For other models all calculations can be found in the Supporting Information.

With all this information in hand, the association constants corresponding to different models can be determined by fitting the dependence of the height of the 3530 cm^{-1} peak on the total concentration of the solution. In practice, the inverse dependence $n(x)$ will be fitted for numerical convenience.

The fitting result for model 1 with the free molecules assumption is shown in Figure 5, and the corresponding result with the free groups assumption is shown in Figure 6. The quality of the fit is visibly better with the free molecules assumption, and this point will be discussed in more detail below.

The results of fitting IR data at 22°C are presented in Table III. The first two columns specify the number of the model and peak attribution assumption respectively. In the third column, the values of the association constants obtained as fitting parameters are shown. In the fourth column, the values of the dimensionless constants are given, which were calculated as $K' = K/v$ where $v = 0.06291/\text{mol}$ is the molar volume of acrylamide based on its density²⁰. The final column gives the values of the AICc parameter³³, which characterizes the quality of the

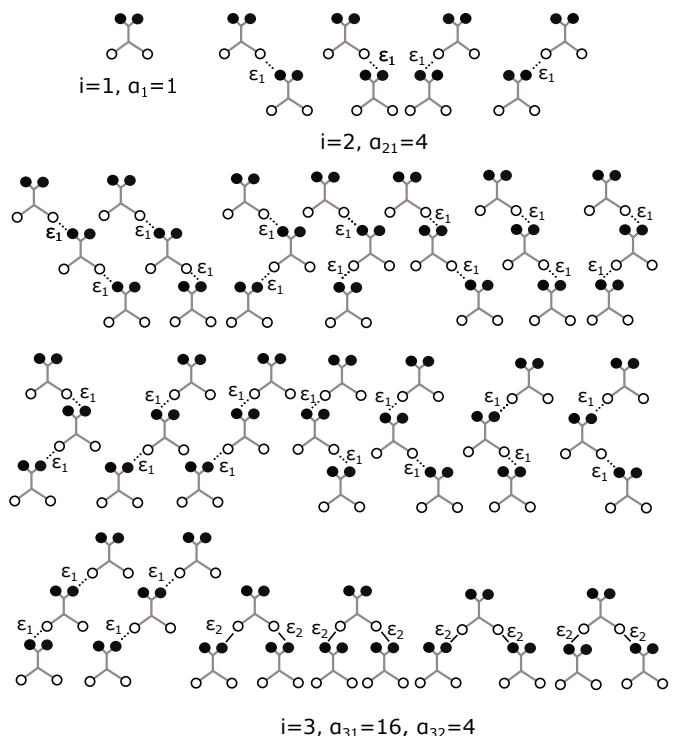


FIG. 4: Schematic representation of aggregates with size up to $i = 3$ in the extension of model 1 when we assume that the energy of the bond is ϵ_1 in the case when a second proton in a donor molecule is free and ϵ_2 otherwise.

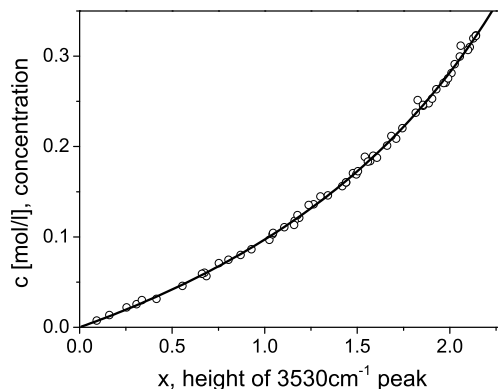


FIG. 5: Fitting of the dependence of the total concentration on the height of the 3530 cm^{-1} IR band at $T = 22^\circ\text{C}$ with model 1 with the free molecules assumption.

non-linear fit^{35,36}. It can be seen that approximately half of the models give the same quality of fit, so based on fitting results exclusively it cannot be said which model is better. However, attributing the peak to free molecules gives a better quality of fit than attributing it to free groups.

Next, the bond energies are determined from the association constants. In order to do this, we perform our

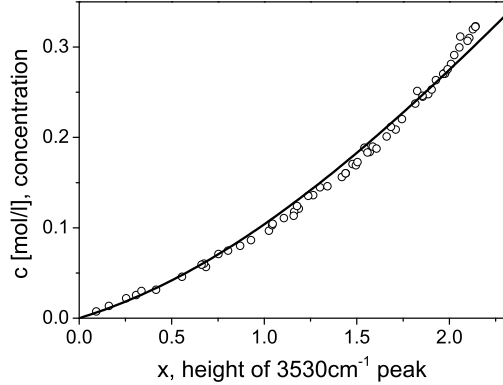


FIG. 6: Fitting of the dependence of the total concentration on the height of the 3530 cm^{-1} IR band at $T = 22^\circ\text{C}$ with model 1 with free groups assumption.

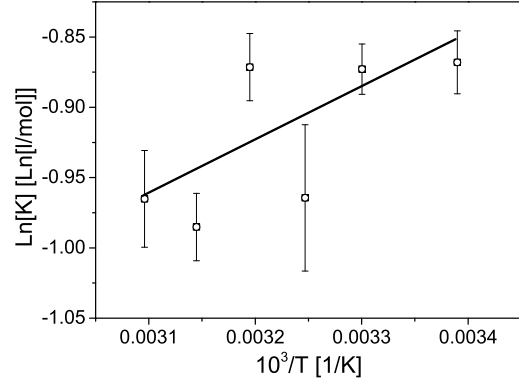


FIG. 7: Dependence of $\ln K$ on $1/T$ for model 1 with the free molecules assumption.

TABLE III: IR data fitting results with models with one association constant.

Model	Peak attr.	K [l/mol] at 22°C	K' at 22°C	AICc at 22°C
0	m ^a	2.4	38.2	-457.6
0	g ^b	1.36	21.6	-298
1	m	0.42	6.68	-503.8
1	g	2.37	37.68	-388.9
1	s ^c	0.65	10.30	-503.9
2	m	0.334	5.31	-502.5
2	g	1.06	16.82	-499.7
2	s	0.442	7.03	-494.2
3	m	0.42	6.68	-499.7
3	g	4.8	76.3	-457.6
3	s	1.02	16.22	-499.3
4	m	0.42	6.68	-503.8
4	g	1.332	21.18	-495.8
4	s	0.65	10.30	-503.9

^a free molecules assumption ^b free groups assumption ^c assumption that 3530 cm^{-1} peak corresponds to free NH_2 groups in unimers and dimers.

measurements and fitting procedures at several different temperatures. Since $\ln K = \ln C - \frac{\epsilon}{kT}$ by definition, ϵ can be determined from a plot of $\ln K$ against $1/T$. It is noteworthy that this procedure can serve as an additional test for the model, because if the model fits the data well then $\ln K (1/T)$ is a linear function with a negative value of the energy of bond formation, $\epsilon < 0$. Examples of such plots are shown in Figures 7 and 8 for model 1 with the free molecules and free groups assumptions. The corresponding quantities C' and ϵ' for the dimensionless association constant K' are shown in Table IV. The first two columns in the table specify the model and peak attribution assumption. The third and fourth columns give the

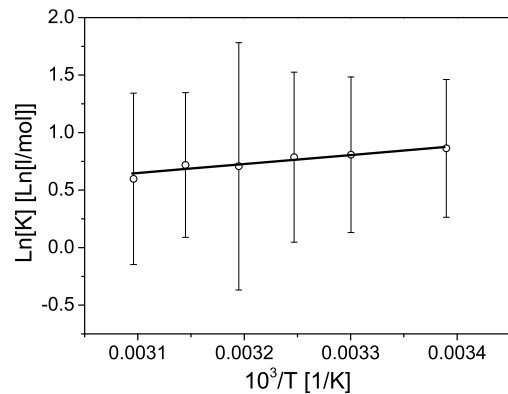


FIG. 8: Dependence of $\ln K$ on $1/T$ for model 1 with the free groups assumption.

values of $\ln C'$ and ϵ respectively, and the values of the coefficients of determination characterizing the quality of the fit of $\ln K'$ vs. $1/T$ are listed in the last column.

It can be seen that the absolute values of the predicted bond energies are close to 1 kcal/mol for all models. However, according to DFT calculations, the absolute value of the bond energy in an acrylamide dimer in a vacuum can be estimated as $7 - 9$ kcal/mol^{13,14}, which is much larger than the values obtained for one-parameter models.

In addition, one can see that, in the limit of infinite temperature, $T \rightarrow \infty$, we have that $K' \rightarrow C'$. Since the C' values are relatively large (see Table IV), $K' > 1$ for all models even at infinite temperatures. This contradicts our intuitive expectation that the strength of hydrogen bonding substantially decreases as the temperature increases. In one of the following sections, models with two association constants are developed, which give a better quality of fit to the data, allowing to obtain a dimer bond energy closer to DFT predictions, and yield physically reasonable asymptotic behavior at high tem-

TABLE IV: Temperature dependence of (dimensionless) association constants and bond energies for the 'one constant' models.

Model	Peak attr.	$\ln C'$	ϵ , [kcal/mol]	r^2
0	m ^a	1.92	-1.03	0.07
0	g ^b	1.12	-1.15	0.75
1	m	0.63	-0.75	0.56
1	g	1.00	-1.55	0.9
1	s	1.01	-0.78	0.66
2	m	0.31	-0.8	0.79
2	g	1.61	-0.72	0.38
2	s	0.50	-0.86	0.92
3	m	0.92	-0.72	0.38
3	g	3.31	-1.03	0.07
3	s	1.57	-0.72	0.37
4	m	0.63	-0.75	0.56
4	g	1.92	-0.675	0.28
4	s	1.01	-0.78	0.66

^a free molecules assumption ^b free groups assumption ^c assumption that 3530cm^{-1} peak corresponds to free NH_2 groups in unimers and dimers.

peratures.

IV. PROPERTIES OF MODELS WITH ONE ASSOCIATION CONSTANT

In the previous section, the concentrations of aggregates of all sizes were determined in models with one association constant. This allows one to study some properties of the models, such as the dependence of the total number of bonds and average size of the aggregates on the total concentration of the solution and on the value of the association constant.

First, let us look at the dependence of the ratio of the concentration of bonds to the concentration of molecules m/c on the value of the association constant at fixed total concentration c , which is shown in Figure 9.

It can be seen that, for model 0, as K increases, m/c tends to a value of 0.5 in accordance with the assumption that in this model only dimers can form. In the case of models 1, 3, and 4, the ratio m/c tends to 1 in the limit of infinite K . This means that, in these models, the number of bonds is always less than the number of molecules in the system. This is explained by the fact that in all of these models molecules either have one acceptor site or one donor site. In contrast, in model 2 there are two bonding sites of each type for each molecule. As a result, when $K > 1/c$, the number of bonds becomes larger than the number of molecules. In addition, since in an aggregate of size i without cycles the number of bonds is always $i - 1$, one can immediately conclude that the assumption about the absence of cycles is wrong when applied to model 2 with $K > 1/c$.

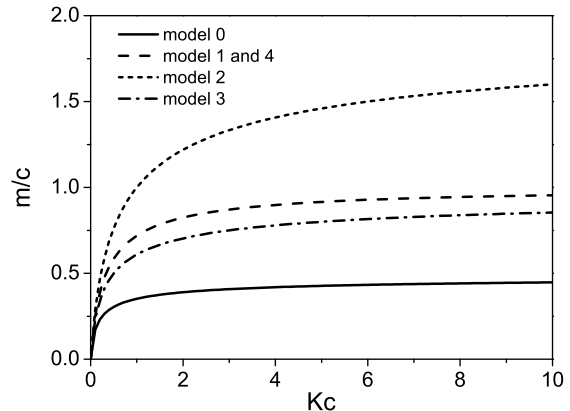


FIG. 9: Dependence of m/c on Kc at fixed total concentration for models with one association constant.

In fact, the restrictions on the applicability of our assumption of the absence of cycles in model 2 are even stronger, because according to our calculations for this model, the total concentration expressed as a function of concentration of unimers can be written in terms of a hypergeometric function as

$$\frac{c}{c_1} = {}_3F_2 \left(\left[\frac{4}{3}, \frac{5}{3}, 2 \right], \left[\frac{5}{2}, 3 \right], \frac{27Kc_1}{2} \right). \quad (18)$$

The right-hand side of Equation 18 is defined only for $|\frac{27Kc_1}{2}| \leq 1$ and is an increasing function of its argument. The first of these facts means that we must have $c_1 \leq \frac{2}{27K}$, and the second means that we must also have $c_1 \geq c / {}_3F_2 \left(\left[\frac{4}{3}, \frac{5}{3}, 2 \right], \left[\frac{5}{2}, 3 \right], 1 \right)$. So, when $K > 2 {}_3F_2 \left(\left[\frac{4}{3}, \frac{5}{3}, 2 \right], \left[\frac{5}{2}, 3 \right], 1 \right) / 27c = 3/8c$, Equation 18 has no solutions. The ratio of m and c at the maximum value of K is $m/c = 2/3$, so the assumption about the absence of cycles fails when the number of bonds per molecule becomes larger than $2/3$.

Another interesting property is the dependence of the average aggregate size on the values of the association constant and concentration. The average aggregation number can be calculated as

$$\langle i \rangle = \frac{\sum_{i=1}^{\infty} i c_i}{\sum_{i=1}^{\infty} c_i}. \quad (19)$$

In Figure 10, the dependence of the average aggregation number on the value of the dimensionless association constant with the volume fraction of acrylamide fixed to the largest experimental volume fraction, $\phi = 0.022$, is shown for models 0, 1, 3 and 4. Figure 11 shows the dependence of the average aggregation number on the volume fraction of acrylamide at association constants obtained from IR measurements for models 0m, 1m, 3m, and 4g. For all models, $\langle i \rangle$ monotonically increases as K' and ϕ increase, and, as expected, $\langle i \rangle$ grows more quickly for models 1 and 4 than for model 3.

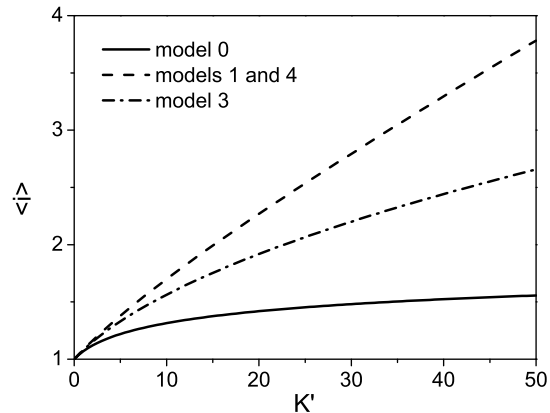


FIG. 10: Dependence of $\langle i \rangle$ on the value of the dimensionless association constant at volume fraction of acrylamide fixed to the largest experimental volume fraction, $\phi = 0.022$.

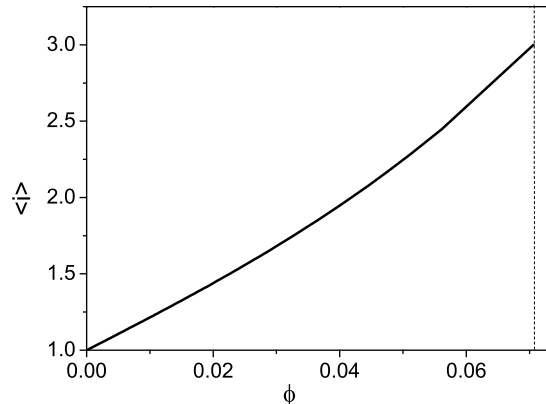


FIG. 12: Dependence of $\langle i \rangle$ on volume fraction of acrylamide with association constant determined from fit of IR data at $T = 22^\circ\text{C}$ for model 2m.

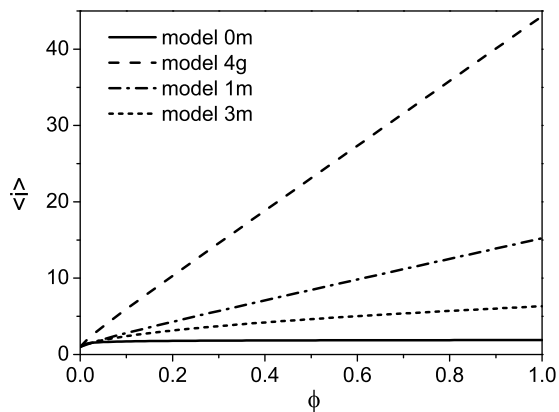


FIG. 11: Dependence of $\langle i \rangle$ on the volume fraction of acrylamide with association constant determined from fitting IR data at $T = 22^\circ\text{C}$.

In the case of model 2, at a fixed value of the concentration there is a maximum value of the association constant above which our assumption about the absence of cycles does not work. Correspondingly, for each value of the association constant, there is a maximum concentration of acrylamide above which model 2 again is not applicable. Within the regime where the model is valid, the average aggregation number is an increasing function of concentration and the association constant and reaches its maximum value of 3 when $Kc_1 = 2/27$. This behavior is illustrated in Figure 12, which shows the dependence of the average aggregation number in model 2 on the volume fraction of acrylamide at a fixed value of the association constant.

V. MODELS WITH TWO ASSOCIATION CONSTANTS

In this section, we consider models in which association is characterized by two association constants. As in the case of alcohols, these constants depend on the bonding state of the other groups belonging to the molecule forming a given hydrogen bond. Again, it is assumed that cycles cannot form. Here only two-constant extensions of models 1, 3 and 4 are constructed. Model 2 has not been considered here, as this case requires significant additional study.

The list of models with two association constants is presented in Table V and the rules for how the energy of a bond depends on its location are illustrated in Figure 13.

All of these cases can be treated analytically and calculations can be found in the Supporting Information. Here, just tables (see Tables VI and VII) of the values of the association constants for these models are shown.

The first thing to note is that the quality of fitting increases as we turn to models with two association constants. Based on the combination of AICc and r_i^2 for the fits of $c(x)$ and $\ln K_i(1/T)$, we can conclude that models 6s (and the equivalent model 10s), 8s and 10g are good. All of these models give the value of the bond energy in a dimer as about 6 – 8 kcal/mol, which corresponds much more closely to the values calculated by DFT^{13,14} discussed earlier. Another important property of these models is that the values they give for both C_1 and C_2 are smaller than those found for one-constant models. So, in the limit of infinite temperature, the association constants K_1 and K_2 will have smaller values.

Fitting also shows that model 5, which includes the assumption that bonding energy is fixed by the bonding state of the neighbor hydrogen in the NH_2 group, is poor. It means that either the bonding state of one hydrogen does not affect the bonding of its neighbor hydro-

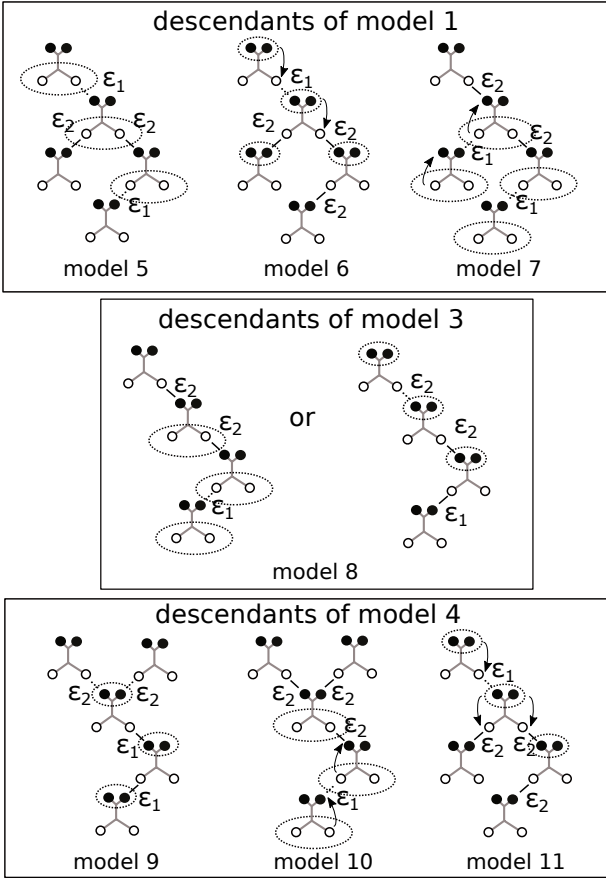


FIG. 13: Two-parameter models.

gen (implying that model 5 reduces to model 1) or that the bonding of one of the hydrogens in the NH_2 group leads to the complete loss of the donor properties of the second hydrogen (so that model 5 reduces to model 3). The same observation applies to the hydrogen bonding sites on oxygen in model 9, which reduces either to model 4 or to model 3.

Similar conclusions can be made with regard to the pair of models 7 and 11, because they show worse fitting results than models 6, 8 and 10. As a result, it can be concluded that the assumption that there are only one or two bonds with energy ϵ_1 at the “top” of each aggregate and all other bonds have energy ϵ_2 is the most probable one according to the fitting results.

It is interesting to note that, in all cases, the value of ϵ_2 is smaller than the value of ϵ_1 , so the formation of the initial dimer is a more energetically favorable process than the addition of subsequent acrylamide molecules to the aggregate. It is also interesting to note that this difference is much larger than in the similar situations in alcohols².

We also can see that there is a contradiction with DFT calculations on formamide and acetamide linear clusters, which predict that the energy per bond increases with the growth of the size of the linear cluster^{16,37}. However, as

TABLE V: List of models with two association constants and no cycles.

Model	Association rules
5	one bond per oxygen, two bonds per NH_2 group, energy of hydrogen bond is determined by the bonding state of the neighbor hydrogen in NH_2 group of the donor molecule
6	one bond per oxygen, two bonds per NH_2 group, energy of hydrogen bond is determined by the bonding state of acceptor in donor molecule
7	one bond per oxygen, two bonds per NH_2 group, energy of hydrogen bond is determined by the bonding state of the NH_2 group in acceptor molecule
8	one bond per oxygen, one bond per NH_2 group, bond energy is determined by bonding state of acceptor in donor molecule or bond energy is determined by bonding state of donor in acceptor molecule (both definitions give equivalent results)
9	two bonds per oxygen, one bond per NH_2 group, bond energy depends on bonding state of acceptor in acceptor molecule
10	two bonds per oxygen, one bond per NH_2 group, bond energy depends on bonding state of NH_2 group in acceptor molecule
11	two bonds per oxygen, one bond per NH_2 group, bond energy depends on bonding state of oxygen group in donor molecule

both approaches involve their own approximations, additional study is needed to understand the reasons for this discrepancy.

VI. STRUCTURE FACTOR OF DIBLOCK COPOLYMER WITH HYDROGEN BONDING BLOCK IN A DISORDERED STATE

Our motivation for developing a theory of hydrogen bonding in acrylamide is to use this model in a self-consistent field theory (SCFT) of block copolymers with hydrogen bonds. To take a first step in this direction, we calculate the structure factor of the disordered state of a diblock copolymer (of degree of polymerization N) with one hydrogen bonding block and one non-hydrogen bonding block in the random-phase approximation (RPA)³⁸.

For the case of alcohols, it was shown by Painter and Coleman that the association constants measured for a monomer can be used to describe association in polymer systems. In order to do this, the association constants should be rescaled to the molar volume of a polymer segment, so that

$$K_i^{\text{polymer}} = \frac{v_{\text{acrylamide}}}{v_{\text{segment}}} K_i^{\text{acrylamide}}. \quad (20)$$

TABLE VI: IR data fitting results with models with two association constant at 22°C.

Model	Peak attr.	K_1 [l/mol]	K'_1	K_2 [l/mol]	K'_2	AICc
5	m	0.38	6.04	0.49	7.79	-501.7
5	g	no good convergence				
5	s	0.69	11.0	0.61	9.7	-501.7
6	m	0.38	6.09	0.44	7.0	-501.7
6	g	0.27	4.29	1.05	16.69	-493.9
6	s	0.72	11.45	0.69	10.92	-501.7
7	m	0.38	6.04	0.45	7.15	-501.7
7	g	0.29	4.54	1.24	19.7	-497.4
7	s	0.74	11.8	0.68	10.8	-501.7
8	m	0.35	5.56	0.61	9.7	-501.5
8	g	0.33	5.24	1.36	21.54	-499.5
8	s	0.56	8.9	0.9	14.3	-501.4
9	m	0.38	6.07	0.49	7.73	-501.7
9	g	0.51	8.11	1.38	21.94	-501.3
9	s	0.70	11.05	0.61	9.70	-501.7
10	m	0.38	6.09	0.44	6.93	-501.7
10	g	0.67	10.59	1.01	16.06	-501.1
10	s	0.72	11.49	0.69	10.92	-501.7
11	m	0.38	6.04	0.45	7.09	-501.7
11	g	no good convergence				
11	s	0.74	11.7	0.68	10.81	-501.7

It is worth mentioning that association constants cannot be measured directly for polymers by infrared spectroscopy, for two reasons. The first is that hydrogen bonding polymers are not soluble in “inert” solvents that do not have specific interactions with the polymer. The second reason is that, since the hydrogen-bonding segments are connected by covalent bonds, all hydrogen-bonding segments have hydrogen-bonding neighbors even at infinite dilution, which means that the polymer segments are not randomly mixed. This is a well-known problem and ideas have been proposed to address it in different areas of polymer theory^{39,40}. Painter, Veytsman and Coleman also proposed an approach to this problem for mixtures of hydrogen-bonding homopolymers⁴¹. However, here we use the random mixing approximation as the simplest starting point for a discussion.

According to one of the basic assumptions of the association model approach, the hydrogen-bonding contribution can be isolated from all other contributions to the free energy, so we have $F = F_0 + F_{\text{HB}}$ where F_0 is the free energy of the system (of total volume V) without hydrogen bonds and F_{HB} is the contribution due to hydrogen bonding. For a melt of diblock copolymer chains where the local and mean volume fractions of the hydrogen-bonding block are $\phi_B(\vec{r})$ and $\langle \phi_B \rangle = f$ respectively, this

TABLE VII: Temperature dependence of (dimensionless) association constants and bond energies for models with two association constants.

Model	Peak attr.	$\ln C'_1$	ϵ_1 [kcal/mol]	r_1^2	$\ln C'_2$	ϵ_2 [kcal/mol]	r_2^2
5	m	-16.8	-11	0.76	6.2	2.4	0.73
5	g	no good convergence					
5	s	-18.7	-12.5	0.76	4.7	1.3	0.56
6	m	-7.2	-5.2	0.91	2.2	0.1	0.03
6	g	-8.7	-5.9	0.71	-1.0	-2.2	0.90
6	s	-9.1	-6.7	0.91	-0.4	-1.6	0.95
7	m	-8.9	-6.2	0.9	4.1	1.2	0.62
7	g	-9.6	-6.5	0.8	-0.13	-1.8	0.95
7	s	-11.6	-8.2	0.9	1.29	-0.6	0.7
8	m	-9.5	-6.6	0.89	2.85	0.3	0.18
8	g	-10.8	-7.3	0.83	-0.39	-2	0.95
8	s	-11.4	-7.9	0.88	0.37	-1.3	0.93
9	m	-16.8	-11	0.76	6.2	2.4	0.73
9	g	no good convergence at higher temperatures					
9	s	-18.7	-12.5	0.76	4.7	1.3	0.56
10	m	-7.2	-5.2	0.91	2.2	0.1	0.03
10	g	-8.5	-6.3	0.9	-1.41	-2.4	0.95
10	s	-9.1	-6.7	0.91	-0.4	-1.6	0.95
11	m	-8.9	-6.2	0.9	4.1	1.2	0.61
11	g	no good convergence					
11	s	-11.6	-8.2	0.9	1.29	-0.6	0.69

gives

$$\begin{aligned} \frac{NF}{kT\rho_0V} &= -\ln \left[\frac{Q[w_A, w_B]}{V} \right] \\ &+ \frac{1}{V} \int d\vec{r} [\chi N \phi_A \phi_B - w_A \phi_A - w_B \phi_B \\ &- \eta (1 - \phi_A - \phi_B) + N f_{\text{HB}}(\phi_B)], \end{aligned} \quad (21)$$

where $\phi_A(\vec{r})$ is the local volume fraction of the non-hydrogen-bonding block, ρ_0 is the bulk segment density, Q is the single-chain partition function, w_A and w_B are the fields corresponding to the two blocks, and η is a Lagrange multiplier that imposes incompressibility. With this expression in hand, the standard derivation of the scattering function in the disordered state in the RPA^{6,42} can be followed to show that

$$\begin{aligned} S^{-1}(k) &= \frac{g(x, 1)}{N \det(S)} - 2\chi + \frac{d^2 f_{\text{HB}}(f)}{d\phi_B^2} \\ &= S_0^{-1}(k) + \frac{d^2 f_{\text{HB}}(f)}{d\phi_B^2}, \end{aligned} \quad (22)$$

where⁶ $g(x, s) \equiv 2[\exp(-sx) + sx - 1]/x^2$ and $S_0^{-1}(k)$ is the inverse scattering function of the block copolymer without hydrogen bonds.

It can therefore be seen that, in the simplest approximation, the effect of hydrogen bonds results in an in-

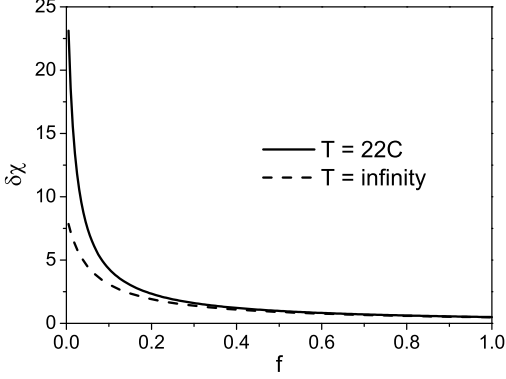


FIG. 14: Dependence of $\delta\chi$ on f for model 1m at $T = 22^\circ\text{C}$ and in the limit of infinite temperature.

crease of the effective Flory-Huggins parameter in this expression (the second derivative of the hydrogen bonding term is negative at all values of composition and association constant) and, furthermore, that the strength of this effect depends on the volume fraction of hydrogen bonding block. This effective χ parameter can then be defined as

$$\chi_{\text{eff}} = \chi - \frac{1}{2} \frac{d^2 f_{\text{HB}}(f)}{d\phi_B^2} = \chi + \delta\chi(f). \quad (23)$$

It is interesting to note that recently it was questioned whether the Flory-Huggins parameter in the case of non-specific interactions does indeed depend on the composition of diblock copolymer, and it was shown in molecular dynamics simulations that χ can be assumed independent of volume fractions^{43,44}. We believe that this result gives additional support to the idea of treating the contributions of non-specific and hydrogen-bonding interactions to the free energy separately.

Now let us turn to the calculations of $\delta\chi$ for a diblock copolymer with a polyacrylamide block based on our models of hydrogen bonding association and the values of association constants deduced from fitting IR data. First, the dependence of $\delta\chi$ on f for models with one association constant is considered. The graph for model 1m is shown in Figure 14. For other models, the plots appear qualitatively the same, so they are not shown here.

The largest absolute value of $\delta\chi$ for one-constant models is attained in the limit as $f \rightarrow 0$ and $\delta\chi$ then monotonically decreases as the fraction of hydrogen-bonding block increases. The decrease of $\delta\chi$ with f is intuitively expected, since, if more neutral segments are mixed in with the network of hydrogen-bonded segments, then more hydrogen bonds need to be broken, and more energy needs to be spent in doing so. It can also be seen that changing the temperature leads to a decrease of the maximal value of $\delta\chi$ at $f \rightarrow 0$, although this value is still very high even at infinite temperature. It is also interesting to note that

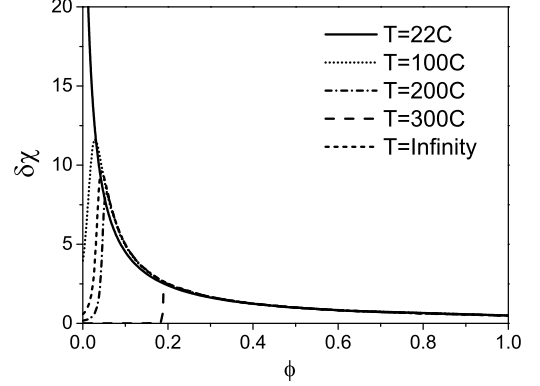


FIG. 15: Dependence of $\delta\chi$ on f for model 6s at different temperatures.

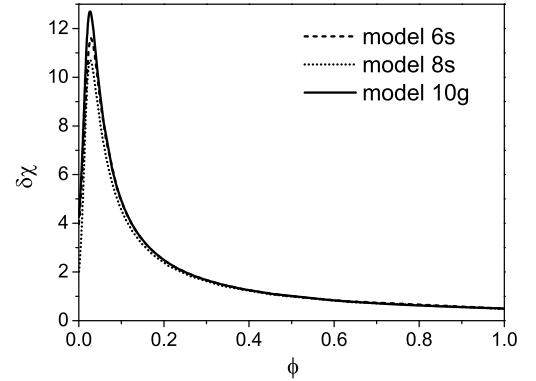


FIG. 16: Dependence of $\delta\chi$ on f for all good models at $T = 100^\circ\text{C}$.

at $f > 0.4$ there is no change in $\delta\chi$ with temperature and that $\delta\chi$ depends only on the volume fraction f .

In the case of models with two association constants, the behavior at small volume fractions is qualitatively different. The dependence of $\delta\chi$ on the volume fraction of the hydrogen-bonding block calculated for model 6s at different temperatures is shown in Figure 15. As the temperature is increased, a peak at a finite value of f appears, which moves to the right as the temperature grows further. However, the behavior of $\delta\chi$ at $f > 0.4$ is similar to the one-parameter models: there is a gradual decrease of $\delta\chi$ as the volume fraction of the hydrogen bonding block is increased and little difference between the curves calculated for different temperatures. It is also interesting to note that the values of $\delta\chi$ for all good two-parameter models are close to each other not only qualitatively, but also quantitatively (see Figure 16).

It can be seen that our current predictions for $\delta\chi$ of polyacrylamide are unrealistically high. This is especially true for small values of f . However, these small values

of f are in fact never reached in polymer systems due to the non-randomness of mixing that we discussed above. In addition, it is important to re-emphasize this analysis as just an initial step on the way to the application of the association model approach to describe hydrogen bonding interactions in block copolymers.

It is also worth mentioning that the common practice of determining χ_{eff} by means of fitting the scattering structure factor in the disordered state in hydrogen-bonding polymers irrespective of the volume fractions probably needs to be changed^{1,45}. Moreover, the values of χ_{eff} determined for hydrogen-bonding block copolymer using this route cannot be used to write the interaction free energy contribution in the form $\chi_{\text{eff}}\phi(1-\phi)$, since the contribution of hydrogen bonds depends on ϕ in a more complicated way. Instead, an appropriate expression provided by the association model approach could be used.

VII. CONCLUSION

In this work, an extension of the association model approach was developed in order to describe the association of molecules with two hydrogen acceptor and two hydrogen donor sites. Models with one association constant were considered, in which it is supposed that all bonds have the same energy, and models with two association constants, in which the bond energy is determined by the local hydrogen-bonding environment.

These models are used to fit FTIR experimental data on solutions of acrylamide in chloroform in order to determine the association constants and their temperature dependence for acrylamide, and found that several models give the same quality of fitting of experimental data. However, models with two association constants in general give better fits than models with one association constant. Moreover, the bond energies in hydrogen bonding dimers for two-constant models are close to the predictions of DFT calculations, which is not the case in one-constant models.

It was also found that, in systems in which two bonds per acceptor site and two bonds per donor site are allowed (such as water), the assumption that cycles are absent ceases to be valid at small concentrations of the hydrogen-bonding substance and there is no non-cyclic solution of the model when the ratio between the number of bonds and number of molecules is larger than 2/3. Interestingly, the largest average aggregation number possible in this model is equal to 3. Based on this result,

one can conclude that in such substances as acrylamide taking into account formation of cycles is essential.

Finally, the structure factor of a disordered state of diblock copolymer with one hydrogen-bonding block and one non-hydrogen-bonding block was calculated in the random phase approximation. We showed that the presence of hydrogen bonds shifts the value of the Flory-Huggins parameter that appears in the expression for the inverse scattering function, and that this shifted value depends on the volume fraction of the hydrogen bonding block, $\chi_{\text{eff}} = \chi + \delta\chi(f)$. The calculations showed that, in general, one-parameter and two-parameter models give similar predictions for the dependence of $\delta\chi$ on the volume fraction of the hydrogen bonding block and on temperature. It is also interesting to note that all good two constant models give very similar quantitative predictions for $\delta\chi$, making the fact that we were unable to determine the best model unimportant from the viewpoint of practical applications.

VIII. EXPERIMENTAL

The FTIR experiments were conducted on a Frontier Perkin Elmer spectrometer. A Specac heatable sealed liquid cell with path length 1 mm and NaCl windows was used. Acrylamide ($\geq 99\%$) and chloroform (anhydrous, stabilized by amylenes, $\geq 99.8\%$) were purchased from Sigma Aldrich.

IX. ACKNOWLEDGEMENT

The work was supported by the Marie Skłodowska-Curie IF “HYBOCOMIX” (ID 704459). The authors thank John Lane for discussions of the model selection criteria, Tom McLeish for discussions of the physics of the association model approach, and Simon Smith, Ivan Ado, Noam Zeilberger and Mark van Hoeij for discussions of the graph theoretical part of the work. The authors also thank Mark van Hoeij for the proof that the solution we found for model 2 is a unique one.

X. SUPPORTING INFORMATION

Supporting Information includes descriptions and calculations for all models and a short reference about AICc criteria for model selection.

* Electronic address: patyukova@gmail.com

† Electronic address: mgreenall@lincoln.ac.uk

¹ J. Kwak, A. K. Mishra, J. Lee, K. S. Lee, C. Choi, S. Maiti, M. Kim, and J. K. Kim, *Macromolecules* **50**, 68136818

(2017).

² M. M. Coleman and P. C. Painter, *Prog. Polym. Sci.* **20**, 1 (1995).

³ A. S. Mahadevi and G. N. Sastry, *Chem. Rev.* **116**, 2775

- (2016).
- ⁴ T. Vasiltsova and A. Heintz, *J. Chem. Phys* **127**, 114501 (2007).
 - ⁵ S. W. Kuo, *J. Polym. Res.* **15**, 459 (2008).
 - ⁶ M. W. Matsen, in *Soft Matter* (Wiley-Blackwell, 2006), chap. 2, pp. 87–178, ISBN 9783527617050.
 - ⁷ A. Dehghan and A. C. Shi, *Macromolecules* **46**, 5796 (2013), ISSN 00249297.
 - ⁸ S. H. Han, J. K. Kim, V. Pryamitsyn, and V. Ganesan, *Macromolecules* **44**, 4970 (2011).
 - ⁹ D. F. Sunday, A. F. Hannon, S. Tein, and R. J. Kline, *Macromolecules* **49**, 48984908 (2016)
 - ¹⁰ Note1, for two cases of alternative approaches see^{46,47}.
 - ¹¹ W.-C. Liao, S. Lilienthal, J. S. Kahn, M. Riutin, Y. S. Sohn, R. Nechushtai, and I. Willner, *Chem. Sci.* **8**, 3362 (2017).
 - ¹² J. Y. Sun, X. Zhao, W. R. Illeperuma, O. Chaudhuri, K. H. Oh, D. J. Mooney, J. J. Vlassak, and Z. Suo, *Nature* **489**, 133136 (2012).
 - ¹³ A. S. Duarte, A. M. Amorim Da Costa, and A. M. Amado, *THEOCHEM* **723**, 63 (2005).
 - ¹⁴ Y.-S. Wang, Y.-D. Lin, and S. D. Chao, *J. Chin. Chem. Soc.* **63**, 968 (2016).
 - ¹⁵ I. Bakó, T. Megyes, S. Bálint, V. Chihaiia, M.-C. Bellissent-Funel, H. Krienke, A. Kopf, and S.-H. Suh, *J. Chem. Phys.* **132**, 014506 (2010).
 - ¹⁶ M. D. Esrafil, H. Behzadi, and N. L. Hadipour, *Theor. Chem. Acc.* **121**, 135 (2008).
 - ¹⁷ A. S. Mahadevi, Y. I. Neela, and G. N. Sastry, *Phys. Chem. Chem. Phys.* **13**, 15211 (2011).
 - ¹⁸ J. M. Stubbs and J. I. Siepmann, *J. Am. Chem. Soc.* **127**, 47224729 (2005).
 - ¹⁹ J. N. Spencer, R. C. Garrett, F. J. Mayer, J. E. Merkle, C. R. Powell, M. T. Tran, and S. K. Berger, *Can. J. Chem.* **58**, 1372 (1980).
 - ²⁰ A. A. Udovenko and L. G. Kolzunova, *J. Struct. Chem.* **49**, 961 (2008).
 - ²¹ R. Böhmer, C. Gainaru, and R. Richert, *Phys. Rep* **545**, 125 (2014), URL <http://dx.doi.org/10.1016/j.physrep.2014.07.005>.
 - ²² A. Milani, C. Castiglioni, E. Di Dedda, S. Radice, G. Canil, A. Di Meo, R. Picozzi, and C. Tonelli, *Polymer* **51**, 2597 (2010), ISSN 00323861, URL <http://dx.doi.org/10.1016/j.polymer.2010.04.002>.
 - ²³ Note2, this estimation is made using Boltzmann distribution and corresponding ground state energies¹⁴.
 - ²⁴ B. A. Veytsman, *J. Phys. Chem.* **94**, 8499 (1990), URL <http://pubs.acs.org/doi/abs/10.1021/j100386a002>.
 - ²⁵ N. N. Coggeshall, E. L. Saier, and N. N. Coggeshall, *J. Am. Chem. Soc.* **73**, 54145418 (1951).
 - ²⁶ M. M. Coleman, X. Yang, P. C. Painter, and J. F. Graf, *Macromolecules* **25**, 4414 (1992).
 - ²⁷ N. G. Mirkin and S. Krimm, *J. Phys. Chem. A* **108**, 54385448 (2004).
 - ²⁸ N. S. Myshakina, Z. Ahmed, and S. A. Asher, *J. Phys. Chem. B* **112**, 1187311877 (2008).
 - ²⁹ J. F. Galan, E. Germany, A. Pawlowski, L. Strickland, and M. G. I. Galinato, *J. Phys. Chem. A* **118**, 53045315 (2014).
 - ³⁰ J.-F. Lu, Z.-Y. Zhou, Q.-Y. Wu, and G. Zhao, *THEOCHEM* **724**, 107 (2005).
 - ³¹ M. D. Esrafil, H. Behzadi, and N. L. Hadipour, *Chem. Phys.* **348**, 175 (2008).
 - ³² Note3, in order to get this solution, we first introduce a function $f(z) = xg(x)$ and get the equation $xf' - 2ff' - 2f = 0$. Then we go to the inverse function $x(f)$ for which $2fx' - x + 2f = 0$. This equation has the solution $x = -2f + A\sqrt{f}$. Finally, we solve the quadratic equation with respect to f .
 - ³³ Note4, the On-Line Encyclopedia of Integer Sequences, published electronically at <https://oeis.org>, May 2018.
 - ³⁴ Note5, smaller values of the AICc parameter correspond to a better quality of fit. Usually, it is assumed that a difference between the AICc parameters of two models of more than 2 is significant and more than 6 is strong. See the supporting information for further details.
 - ³⁵ H. Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle* (Springer New York, New York, NY, 1998), ISBN 978-1-4612-1694-0, URL https://doi.org/10.1007/978-1-4612-1694-0_15.
 - ³⁶ N. Sugiura, *Commun. Stat. Theory Methods* **7**, 13 (1978), URL <https://doi.org/10.1080/03610927808827599>.
 - ³⁷ N. Kobko and J. J. Dannenberg, *J. Phys. Chem. A* **107**, 10389 (2003).
 - ³⁸ L. Leibler, *Macromolecules* **13**, 1602 (1980), ISSN 0024-9297, 0402594v3, URL <http://pubs.acs.org/doi/abs/10.1021/ma60078a047>.
 - ³⁹ D. C. Morse and J. K. Chung, *J. Chem. Phys.* **130**, 224901 (2009), URL <https://doi.org/10.1063/1.3108460>.
 - ⁴⁰ T. P. Lodge and T. C. B. McLeish, *Macromolecules* **33**, 5278 (2000), ISSN 00249297.
 - ⁴¹ P. C. Painter, B. Veytsman, S. Kumar, S. Shenoy, J. F. Graf, Y. Xu, and M. M. Coleman, *Macromolecules* **30**, 932 (1997).
 - ⁴² L. Leibler, *Macromolecules* **13**, 1602 (1980), ISSN 0024-9297, 0402594v3, URL <http://pubs.acs.org/doi/abs/10.1021/ma60078a047>.
 - ⁴³ T. Ghasimakbari and D. C. Morse, *Macromolecules* **51**, 23352348 (2018).
 - ⁴⁴ S. Mogurampelly, B. H. Nguyen, and V. Ganesan, *J. Chem. Phys.* **141**, 244904 (2014).
 - ⁴⁵ D. P. Sweat, M. Kim, A. K. Schmitt, D. V. Perroni, C. G. Fry, M. K. Mahanthappa, and P. Gopalan, *Macromolecules* **47**, 6302 (2014), ISSN 15205835.
 - ⁴⁶ N. Lefèvre, K. C. Daoulas, M. Müller, J. F. Gohy, and C. A. Fustin, *Macromolecules* **43**, 77347743 (2010).
 - ⁴⁷ X. Zhang, J. Lin, L. Wang, L. Zhang, J. Lin, and L. Gao, *Polymer* **78**, 69 (2015).

Supplementary Materials: Hydrogen bonding in acrylamide and its role in the scattering behavior of acrylamide-based block copolymers

XI. MODELS

A. Model 0

In model 0, we assume that the solution contains only monomers and linear dimers (see Figure S1), and that chemical equilibrium with respect to hydrogen bonding association is described by a single association constant K . Then, the total concentration of the solution and the concentration of unimers are related to each other by the equation

$$c = c_1 + 2c_2, \quad (\text{S1})$$

where the concentration of dimers is $c_2 = 4Kc_1^2$. The factor of 4 appears here because there are four ways to form a dimer from two identical molecules (due to the presence of two hydrogens in the NH_2 group and two lone electron pairs on the oxygen).

The free energy of hydrogen bonding of a system of volume V with N molecules and M dimers can be written as

$$F_{\text{HB}} = M\epsilon - kT \ln(p^M \Xi), \quad (\text{S2})$$

where ϵ is the energy of a hydrogen bond, $p = C/V$ (where C is a constant) is the probability that two molecules will meet and orient with respect to each other to form a bond, and Ξ is the combinatorial number of ways to form M dimers out of N molecules, such that

$$\Xi = \frac{N!2^M}{(N-M)!} \frac{(N-M)!2^M}{(N-2M)!} \frac{1}{M!} = \frac{N!4^M}{(N-2M)!M!}, \quad (\text{S3})$$

where the first factor is the number of ways of choosing M acceptor molecules, the second factor is the number of ways of choosing M donor molecules, the factor 2^M takes into account that each molecule has two hydrogens (in the case when we assume that two bonds per oxygen are possible then an additional factor of 2^M appears), and the last factor takes into account the indistinguishability of bonds.

Minimizing the free energy with respect to M yields

$$\frac{M}{4(N-2M)^2} = \frac{K}{V}, \quad (\text{S4})$$

where $K = C \exp(-\epsilon/kT)$, or, in terms of concentrations,

$$\frac{m}{4(c-2m)^2} = K. \quad (\text{S5})$$

As the concentration of unimers in the system is $c_1 = c - 2m = c - 8Kc_1^2$, we get $m = c_2 = 4Kc_1^2$ and $c = c_1 + 8Kc_1^2$ in agreement with Equation S1.

Let us now try to determine the association constant by fitting the dependence of the concentration on the height of the 3530cm^{-1} peak.

1. Model 0m

First, we assume that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in unimers. In this case, $c_1 = Ax$, where x is the height of the peak and A is some constant. Substituting this in Equation S1 gives the fitting equation

$$c = Ax + 8KA^2x^2. \quad (\text{S6})$$

A fit of the dependence of peak intensity on concentration at $T = 22^\circ\text{C}$ is shown in Figure S2. The quality of the nonlinear fit can be quantified by the Akaike Information Criterion, on which further details are given at the end of this document. The value of this quantity for the current fit is $\text{AICc} = -457.6$.

According to the definition of the association constant in our model, $K = C \exp(-\epsilon/kT)$, $\ln K$ should depend linearly on inverse temperature $1/T$, and this dependence, together with a linear weighted fit, is shown in Figure S3. This yields estimates for the model parameters of $\ln C = -0.84 \pm 3 \ln[1/\text{mol}]$ and $\epsilon = -1.03 \pm 1.8 \text{ kcal/mol}$. The poor quality of the fit in this case is reflected in the low value of the coefficient of determination, $r^2 = 0.07$.

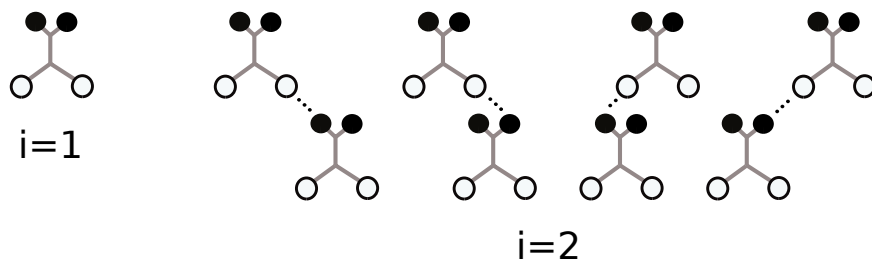


FIG. S1: Schematic representation of the amide group and all four ways of forming a hydrogen bonded linear dimer from identical molecules. Hydrogen bonds are represented by dots. In Model 0, we assume that only unimers and linear dimers are present in the solution.

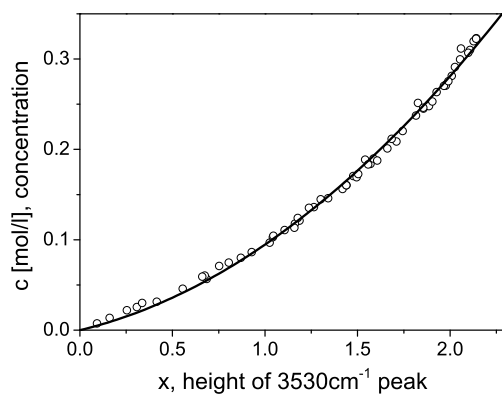


FIG. S2: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 0m.

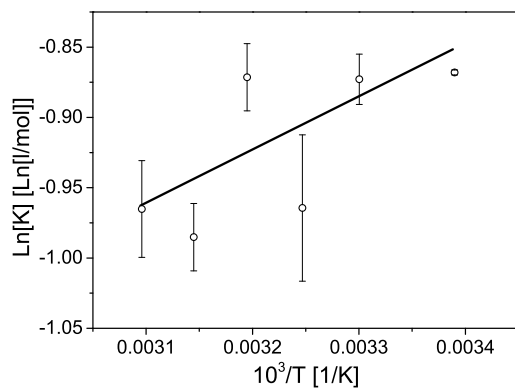


FIG. S3: Dependence of $\ln K$ on $1/T$ for model 0m.

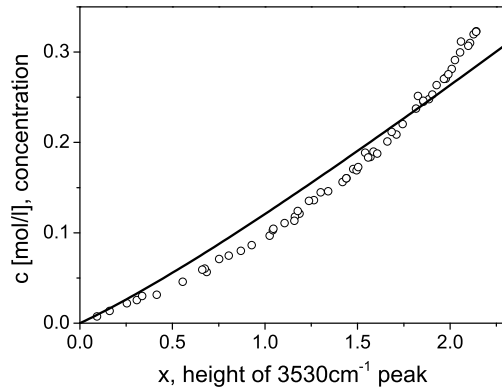


FIG. S4: Fit of the dependence of total concentration on height of 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 0g.

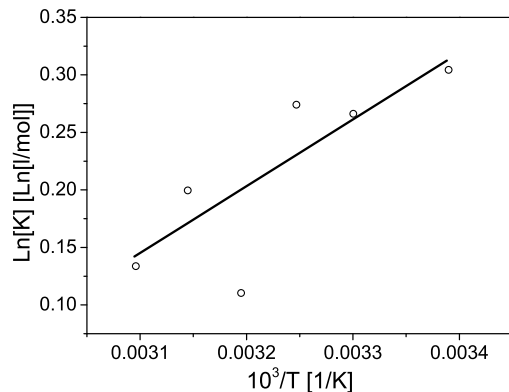


FIG. S5: Dependence of $\ln K$ on $1/T$ for model 0g. The error bars are too large to be shown.

2. Model 0g

It is now assumed that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of free NH_2 groups. In this case, $Ax = c_1 + 4Kc_1^2$, and the fitting equation is

$$c = \frac{1 + 16AKx - \sqrt{1 + 16AKx}}{8K}. \quad (\text{S7})$$

A fit of peak intensity versus concentration at $T = 22^\circ\text{C}$ is shown in Figure S4. This fit is visibly less successful than that for model 0m, and AICc takes the higher value of -298 .

The dependence of $\ln K$ on $1/T$ is shown in Figure S5. The estimates of the model parameters are $\ln C = -1.65 \pm 0.5 \text{ ln[l/mol]}$ and $\epsilon = -1.15 \pm 0.3 \text{ kcal/mol}$. The quality of the fit is better than in model 0m, and this is shown by the higher value of the coefficient of determination, $r^2 = 0.75$. However, we note that this apparent improvement may be offset by the very large error bars on $\ln K$, which probably result from the poor quality nonlinear fit in Figure S4.

The free energy density due to hydrogen bonding in model 0 in terms of the volume fraction of hydrogen bonding molecules $\phi = cv$ and the dimensionless association constant $K' = K/v$ has the form

$$f_{\text{HB}} = \frac{(\sqrt{1 + 32K'\phi} - 1)^2}{64K'} + \phi \ln \frac{\sqrt{1 + 32K'\phi} - 1}{16K'\phi}. \quad (\text{S8})$$

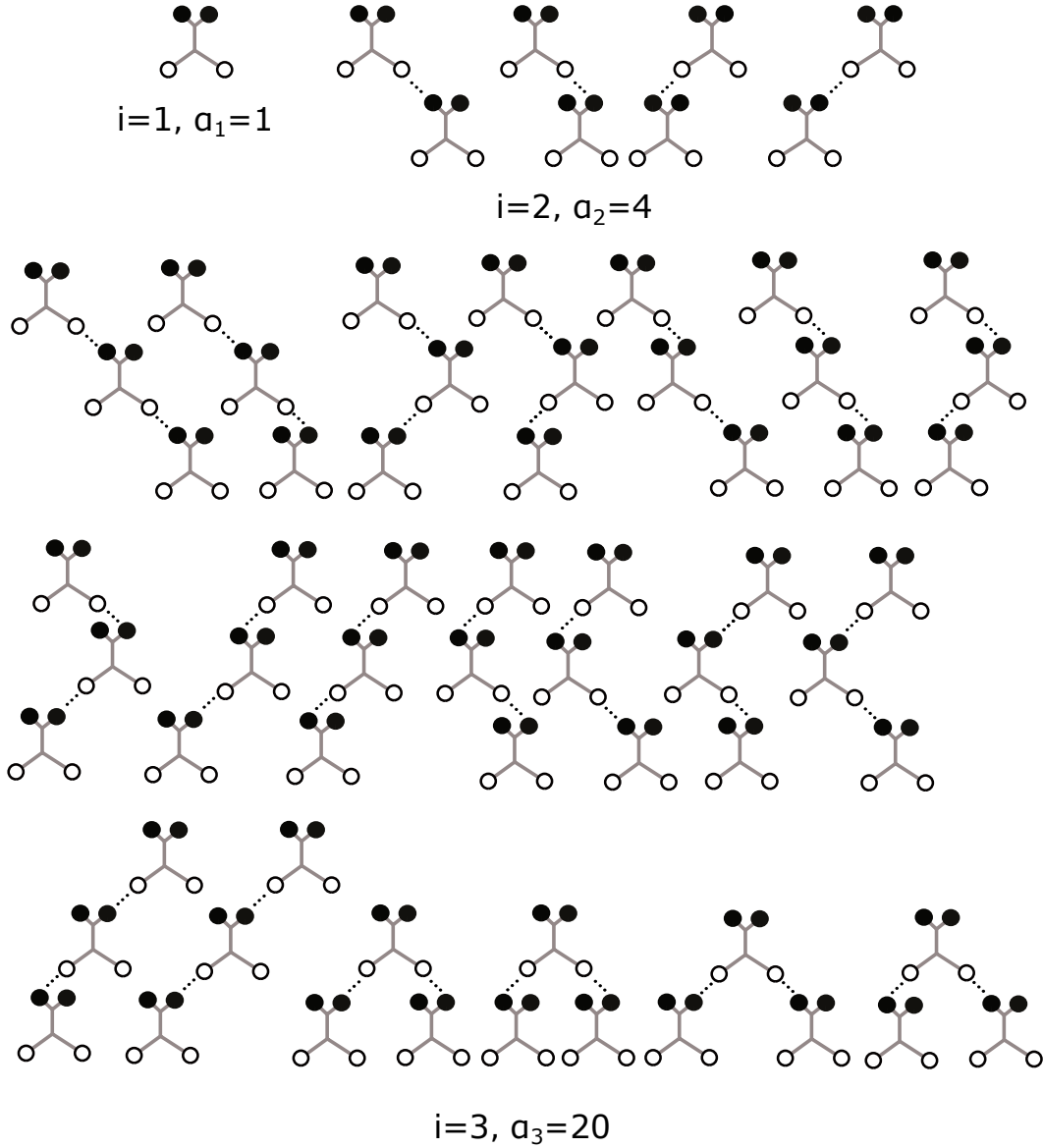


FIG. S6: Schematic representation of possible aggregates with size up to $i = 3$ in model 1.

B. Model 1

In model 1, we assume that we have one bond per oxygen, two bonds per NH_2 group, one association constant and no cycles. This means that the aggregates are tree-shaped (see Figure S6).

The free energy of hydrogen bonding can be written as

$$F = M\epsilon - kT \ln(p^M \Xi), \quad (\text{S9})$$

where M is the number of hydrogen bonds, p is the probability that a donor and an acceptor form a bond, and Ξ is the number of ways to form M bonds, given in this case by

$$\Xi = \frac{N! 2^M}{(N-M)! (2N-M)! M!}, \quad (\text{S10})$$

where the first factor is the number of ways to choose an acceptor, the second is the number of ways to choose a donor, and the final factor takes into account the fact that all bonds are identical. Substituting Equation (S10) into

Equation (S9) and using Stirling's formula gives

$$\frac{F}{kT} = M \left(\frac{\epsilon}{kT} - \ln p \right) + N \ln \frac{(N-M)(2N-M)^2}{4N^3} + M \ln \frac{Me}{(N-M)(2N-M)}. \quad (\text{S11})$$

After minimization with respect to M , we find, in terms of concentrations $m = M/V$ and $c = N/V$,

$$\frac{m}{2(c-m)(2c-m)} = K. \quad (\text{S12})$$

Let us suppose that the concentration of aggregates of size i can be expressed as

$$c_i = \alpha_i K^{i-1} c_1^i, \quad (\text{S13})$$

where the α_i are unknown coefficients. Then, for the concentration of bonds m and total concentration c we have

$$m = \sum_{i=1}^{\infty} (i-1) c_i \quad (\text{S14})$$

and

$$c = \sum_{i=1}^{\infty} i c_i. \quad (\text{S15})$$

In order to find α_i , we substitute expressions S14 and S15 into Equation S12 and equate coefficients in front of like powers of c_1 . Using this method, we can calculate the values of the coefficients, which in this case are (starting from $i = 1$) 1, 4, 20, 672, 4224, 27456, ... Using the On-Line Encyclopedia of Integer Sequences (published electronically at <https://oeis.org>, May 2018), we can assume that the general formula for a term of this sequence most probably has the form (in the main paper a more elegant way to get this result is described)

$$\alpha_i = 2^{i-1} \frac{(2i)!}{(i+1)!i!}. \quad (\text{S16})$$

The sequence $\beta_i = (2i)!/(i+1)!i!$ is known as the Catalan numbers. It is known that these numbers represent the number of different rooted binary trees with $i+1$ leaves. In our case, we have an additional factor of 2^{i-1} , since each molecule apart from the root can be added in two ways to form a bond with one of the free hydrogens because there are two bonding sites on the oxygen. All aggregates allowed in this model with size up to $i = 3$ are shown in Figure S6. So we can say that the physical meaning of α_i is the number of ways to form an aggregate of size i out of i molecules.

It is also interesting to note that, by looking at Figure S6, it can be seen that the aggregates can be built recursively from each other, so a generating function $G(z)$ can be written as

$$G(z) = 1 + 4zG(z) + 4z^2G(z)^2, \quad (\text{S17})$$

where $z = Kc_1$ is the multiplicative factor that is introduced when the size of the aggregate is increased by one. Solution of this equation gives

$$G(z) = \frac{1 - 4z - \sqrt{1 - 8z}}{8z^2}, \quad (\text{S18})$$

and this generating function can be expanded with respect to z to give the values of α_i .

With this expression for α_i in hand, the total concentration of the solution can be calculated as

$$c = \frac{1 - 4Kc_1 - \sqrt{1 - 8Kc_1}}{8K^2c_1\sqrt{1 - 8Kc_1}}. \quad (\text{S19})$$

1. Model 1m

Let us assume first that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in unimers. In this case, $c_1 = Ax$, where x is the height of the peak and A is some constant. Substituting this in Equation S19 gives the fitting equation:

$$c = \frac{1 - 4KAx - \sqrt{1 - 8KAx}}{8K^2Ax\sqrt{1 - 8KAx}}. \quad (\text{S20})$$

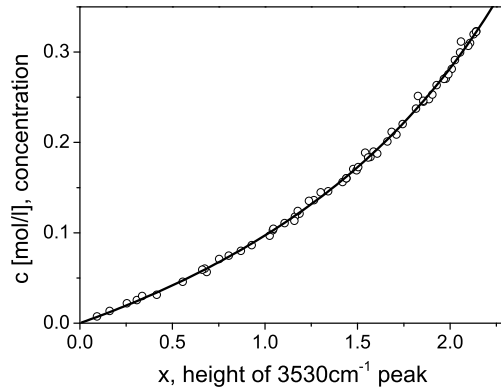


FIG. S7: Fit of the dependence of the total concentration on the height of 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 1m.

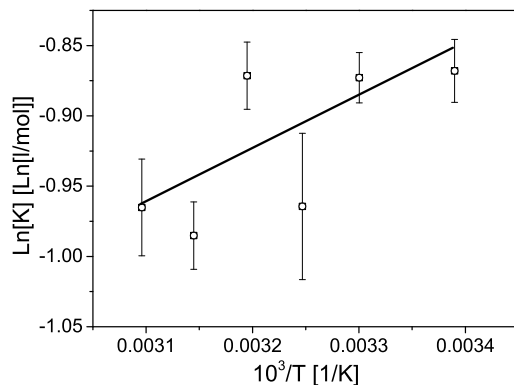


FIG. S8: Dependence of $\ln K$ on $1/T$ for model 1m.

The results of this fit are shown in Figures S7 and S8. The value of AICc for the fit in Figure S7 is -503.8 . The dependence of $\ln K$ on $1/T$ is shown in Figure S8, and this fit gives estimates of the model parameters of $\ln C = -2.14 \pm 0.6 \ln[\text{l/mol}]$ and $\epsilon = -0.79 \pm 0.2 \text{ kcal/mol}$, with a coefficient of determination of $r^2 = 0.56$.

2. Model 1g

In model 1g, it is assumed that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of NH_2 free groups both in free molecules and in aggregates. Therefore, in order to find a fitting equation, we need to calculate the concentration of free groups. However, it turns out that it is impossible to do this in the framework of model 1 because the number of free groups depends on the structure of the aggregate. However, model 5 (to be introduced later) does include the relevant information about the structure of the aggregate, and reduces to model 1 when its two association constants are set equal to each other. Therefore, the calculation of the fitting expression is carried out in model 5, and K_1 and K_2 are both set equal to K at the end. This gives

$$c = \frac{1 - 4AKx + 16A^2K^2x^2 - (1 - 4AKx)\sqrt{1 + 16A^2K^2x^2}}{8AK^2x}. \quad (\text{S21})$$

The results of the fit are shown in Figures S9 and S10. The quality of fit in Figure S9 can be characterized by the parameter AICc = -388.9 , and the fit shown in Figure S10 estimates the model parameters to be $\ln C = -1.77 \pm 0.4 \ln[\text{l/mol}]$ and $\epsilon = -1.55 \pm 0.3 \text{ kcal/mol}$, with $r^2 = 0.9$.

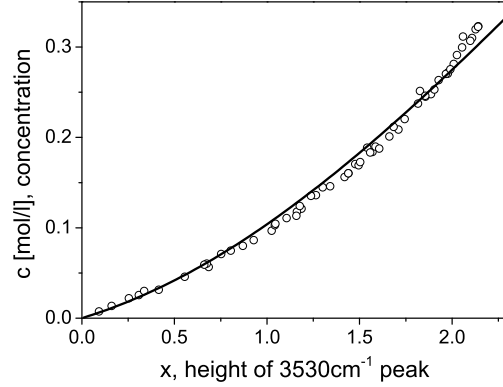


FIG. S9: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 1g.

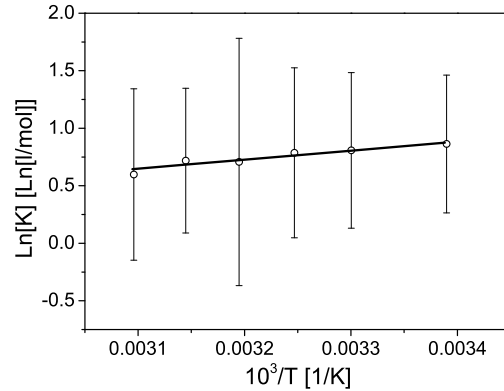


FIG. S10: Dependence of $\ln K$ on $1/T$ for model 1g.

3. Model 1s

Here we assume that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of NH_2 free groups in free molecules and dimers. In this case, the relation between c_1 and x takes the form $Ax = c_1 + 4Kc_1^2$.

The fitting results are shown in Figures S11 and S12. The value of AICc for the fit in Figure S11 is -503.9 , and the estimates for the model parameters corresponding to Figure S12 are $\ln C = -1.75 \pm 0.5 \ln[\text{l/mol}]$ and $\epsilon = -0.78 \pm 0.3 \text{ kcal/mol}$, with $r^2 = 0.66$.

The free energy density of hydrogen bonding in model 1 in terms of the volume fraction of hydrogen bonding molecules $\phi = cv$ and dimensionless association constant $K' = K/v$ has the form

$$f_{\text{HB}} = m + \phi \ln \frac{(\phi - m)(2\phi - m)^2}{4\phi^3}, \quad (\text{S22})$$

where m is a solution of the equation $m/[2(\phi - m)(2\phi - m)] = K'$.

C. Model 2

In model 2, the following assumptions are made: oxygen can form two bonds, the NH_2 group can form two bonds, there are no cycles and there is one association constant. The range of allowed aggregates for this case with sizes up to $i = 3$ is shown in Figure S13.

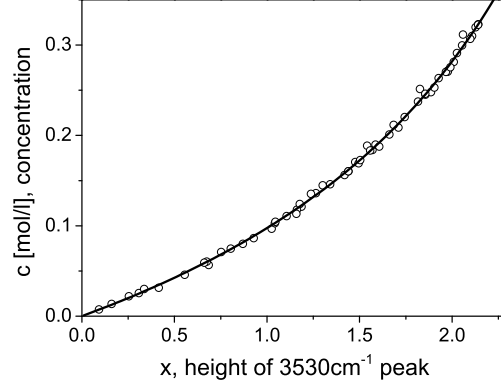


FIG. S11: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 1s.

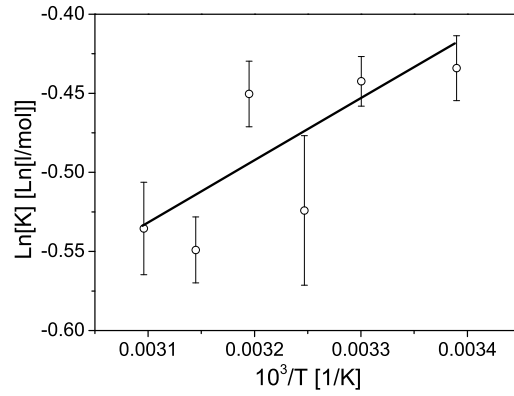


FIG. S12: Dependence of $\ln K$ on $1/T$ for model 1s.

The free energy of hydrogen bonding can be written as

$$F = M\epsilon - kT \ln(p^M \Xi), \quad (\text{S23})$$

where M is the number of hydrogen bonds and the number of ways to form these bonds is

$$\Xi = \left(\frac{2N!}{(2N - M)!} \right)^2 \frac{1}{M!}. \quad (\text{S24})$$

Minimization of the free energy yields

$$\frac{m}{(2n - m)^2} = K. \quad (\text{S25})$$

Next, it is assumed that the concentration of aggregates of size i can be written as

$$c_i = \alpha_i K^{i-1} c_1^i, \quad (\text{S26})$$

so that the total concentration is

$$n = \sum_{i=1}^{\infty} i c_i \quad (\text{S27})$$

and the concentration of bonds is

$$m = \sum_{n=1}^{\infty} (n - 1) c_n. \quad (\text{S28})$$

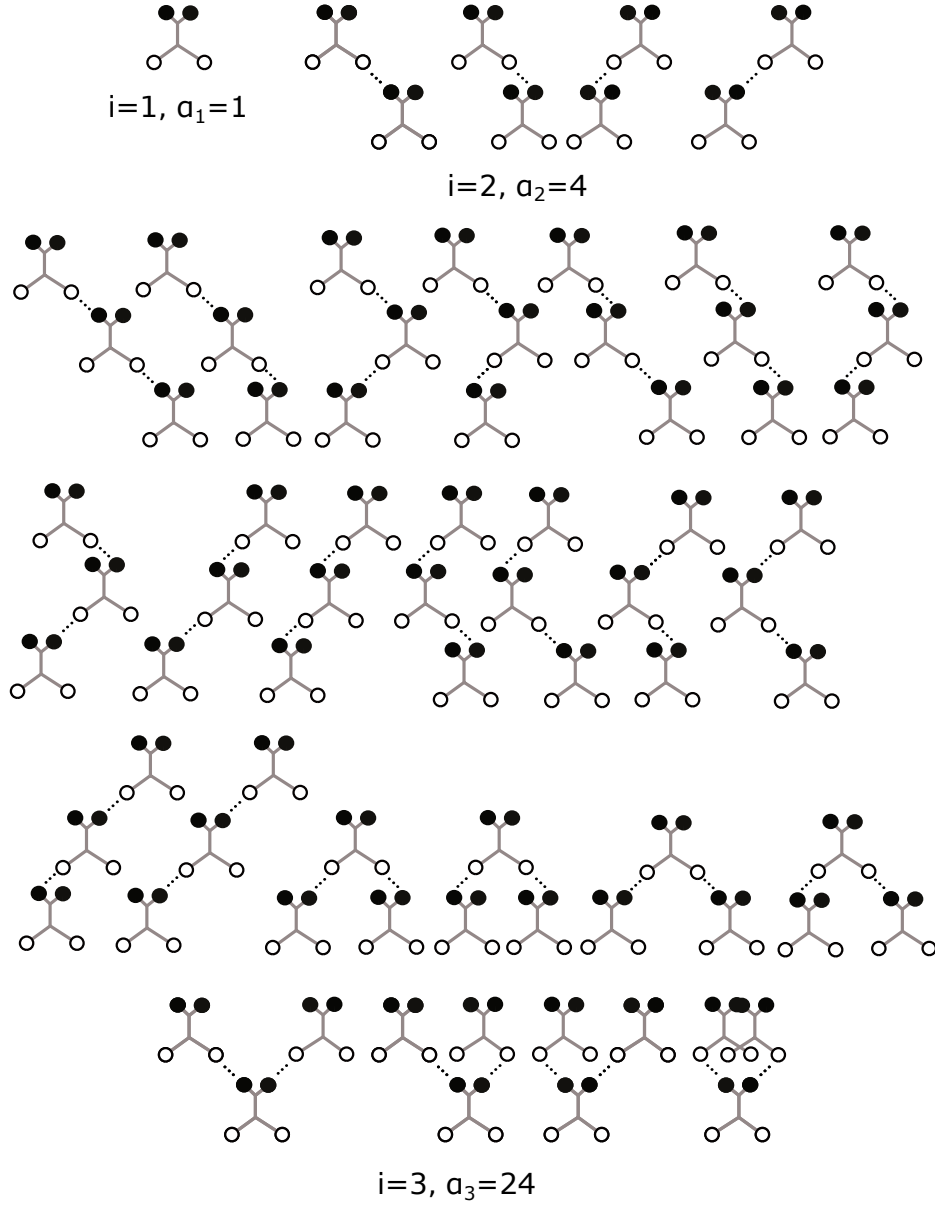


FIG. S13: Schematic representation of possible aggregates with size up to $i = 3$ in model 2.

Then, we substitute Equations S27 and S28 into Equation S25 written in the form $m = K(2n - m)^2$ and, equating coefficients in front of like powers of c_1 , we can calculate the first terms of the α_i sequence. The OEIS tells us that this sequence is probably that known as A000309, whose i^{th} term is given by

$$\alpha_i = 2^i \frac{(3i)!}{(2i+1)!(i+1)!}. \quad (\text{S29})$$

In the following section, a proof is given that the sequence A000309 is indeed the set of coefficients α_i that satisfy Equation S25. Using this result for the α_i , the total concentration can be written down as

$$n = c_1 \cdot {}_3F_2 \left(\left[\frac{4}{3}, \frac{5}{3}, 2 \right], \left[\frac{5}{2}, 3 \right], \frac{27Kc_1}{2} \right). \quad (\text{S30})$$

1. The proof

This proof was provided by Mark van Hoeij.

Let us consider α_i , m and n defined according to Equations S29, S28, and S27 respectively. Then, let us define a function

$$y = 2n - m = \frac{1}{K} \sum_{i=1}^{\infty} (i+1) \alpha_i x^i, \quad (\text{S31})$$

where $x = Kc_1$. Next, introduce $M = Km$ and $Y = ky$. Then, Equation S25 is equivalent to

$$M = Y^2. \quad (\text{S32})$$

If a function Z is defined by

$$Z = 1 + Y = \sum_{i=0}^{\infty} (i+1) \alpha_i x^i \quad (\text{S33})$$

(the difference with Y is that the summation starts from 0), it can be verified that

$$(27x - 2)xZ'' + (54x - 3)Z' + 6Z = 0. \quad (\text{S34})$$

Indeed, since

$$3Z' = 3 \sum_{i=0}^{\infty} (i+1) i \alpha_i x^{i-1} = 3 \sum_{i=0}^{\infty} (i+1) (i+2) \alpha_{i+1} x^i \quad (\text{S35})$$

and

$$2xZ'' = 2 \sum_{i=0}^{\infty} (i+1) i (i-1) \alpha_i x^{i-1} = 2 \sum_{i=0}^{\infty} i (i+1) (i+2) \alpha_{i+1} x^i, \quad (\text{S36})$$

Equation S34 is equivalent to

$$(27i^2 + 27i + 6) \alpha_i = (i+2)(2i+3) \alpha_{i+1}, \quad (\text{S37})$$

which is true for α_i defined by Equation S29. Differentiating Equation S34 gives

$$(27x - 2)xZ''' + (108x - 5)Z'' + 60Z' = 0, \quad (\text{S38})$$

or, since $Y' = Z'$,

$$(27x - 2)xy'''' + (108x - 5)Y'' + 60Y' = 0, \quad (\text{S39})$$

and it can be found that $u = Y^2$ satisfies the equation

$$(27x - 2)^2 x^3 u'''''' + 2x^2 (270x - 13)(27x - 2)u'''' + 35x(2592x^2 - 246x + 5)u'''' + (204120x^2 - 13440x + 140)u''' + (146160x - 5040)u'' + 20160u' = 0. \quad (\text{S40})$$

To show this, we note that, if we take Y^2 and differentiate repeatedly, then all the resulting expressions can be written in terms of products of Y , Y' and Y'' , because all instances of Y''' can be eliminated with Equation S39. Next, we verify that $M = \sum_{i=1}^{\infty} (i-1) \alpha_i x^i$ also satisfies Equation S40. Since both M and Y^2 satisfy Equation S40, the functions M and Y^2 are equal if the first six terms in their expansions in powers of x coincide (because the differential equation is sixth order), which we can check by direct computation.

2. Model 2m

Let us assume first that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in unimers. In this case, $c_1 = Ax$ where x is the height of the peak and A is some constant. Substituting this in Equation S30 gives

$$n = Ax \cdot {}_3F_2 \left(\left[\frac{4}{3}, \frac{5}{3}, 2 \right], \left[\frac{5}{2}, 3 \right], \frac{27KAx}{2} \right). \quad (\text{S41})$$

The results of the fit are shown in Figures S14 and S15. The value of AICc in Figure S14 is -502.5 , and the estimates of the model parameters given by the fit in Figure S15 are $\ln C = -2.5 \pm 0.3 \ln[\text{l/mol}]$ and $\epsilon = -0.8 \pm 0.2 \text{ kcal/mol}$ with $r^2 = 0.75$.

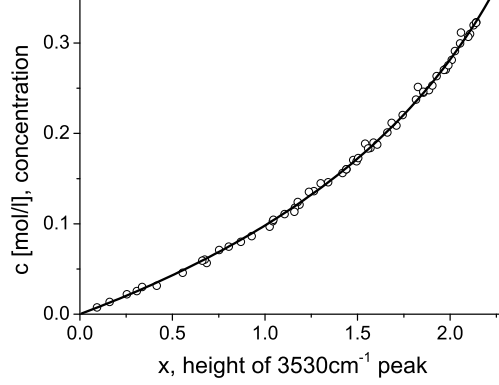


FIG. S14: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 2m.

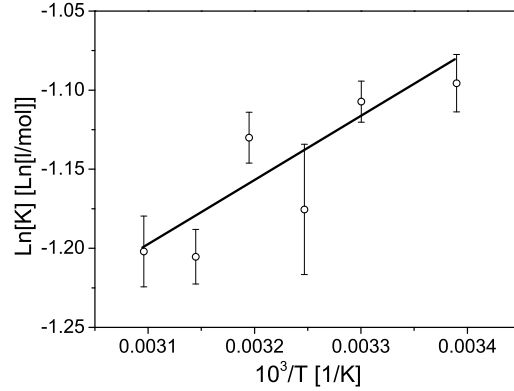


FIG. S15: Dependence of $\ln K$ on $1/T$ for model 2m.

3. Model 2g

Let us assume now that the 3530cm^{-1} peak corresponds to out-of-phase vibrations of NH_2 free groups. As in the case of model 1g, we need to include more information about the aggregate in order to calculate the concentration of free groups. If we distinguish between molecules with both hydrogens bonded (N_2 is the number of such molecules) and only one bonded hydrogen in the amide group (N_1 is the number of such molecules) we can write the number of ways to form bonds as (see model 6 for more details)

$$\begin{aligned} \Xi &= \frac{2N!}{(2N-M)!} \frac{N!2^{N_1}}{(N-N_2-N_1)!N_1!N_2!} \\ &= \frac{2N!}{(2N-M)!} \frac{N!2^{2N-M-2N_f}}{N_f!(2N-M-2N_f)!(N_f+M-N)!} \end{aligned} \quad (\text{S42})$$

where $M = N_1 + 2N_2$ and $N_f = N - N_1 - N_2$. Minimizing the free energy and eliminating M gives

$$c = \frac{n_f}{(1-2Kn_f)^2} = \frac{Ax}{(1-2KAx)^2}, \quad (\text{S43})$$

where, according to our peak attribution assumption, $n_f = Ax$.

The results of this fit are shown in Figures S16 and S17. The value of AICc for the fit in Figure S16 is -499.7 , and the fit in Figure S17 yields estimates for the model parameters of $\ln C = -1.2 \pm 0.7 \ln[\text{l/mol}]$ and $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$, with $r^2 = 0.38$.

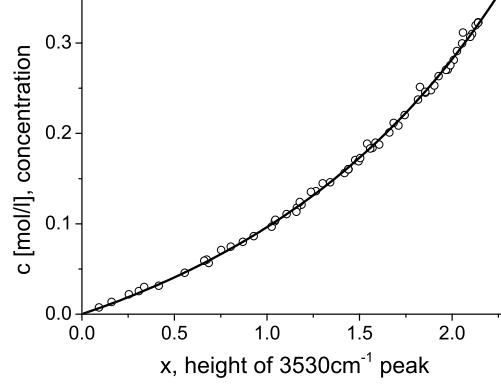


FIG. S16: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 2g.

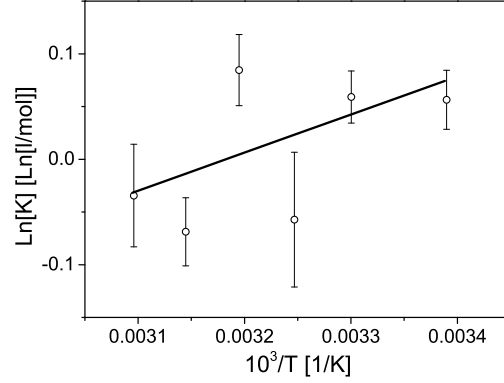


FIG. S17: Dependence of $\ln K$ on $1/T$ for model 2g.

4. Model 2s

Let us assume here that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the free NH_2 group in unimers and dimers. In this case, $Ax = c_1 + 4Kc_1^2$, where x is the height of the peak and A is a constant. Substituting this in Equation S30 yields

$$c = \frac{-1 + \sqrt{1 + 16AKx}}{8K} \cdot {}_3F_2 \left(\left[\frac{4}{3}, \frac{5}{3}, 2 \right], \left[\frac{5}{2}, 3 \right], \frac{27}{16} \left(-1 + \sqrt{1 + 16AKx} \right) \right). \quad (\text{S44})$$

The results of the fit are shown in Figures S18 and S19. The value of AICc for the fit shown in Figure S18 is -494.2 , and the fit in Figure S19 yields estimates for the model parameters of $\ln C = 0.50 \pm 0.2 \ln[1/\text{mol}]$ and $\epsilon = -0.86 \pm 0.1 \text{ kcal/mol}$, with $r^2 = 0.92$.

The free energy due to hydrogen bonding in model 2 in terms of the volume fraction of hydrogen bonding molecules $\phi = cv$ and the dimensionless association constant $K' = K/v$ has the form

$$f_{\text{HB}} = m + 4\phi \ln \frac{(2\phi - m)}{2\phi}, \quad (\text{S45})$$

where m is a solution of the equation $m/(2\phi - m)^2 = K'$.

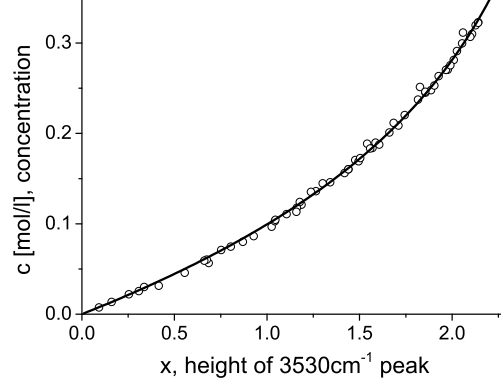


FIG. S18: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 2s.

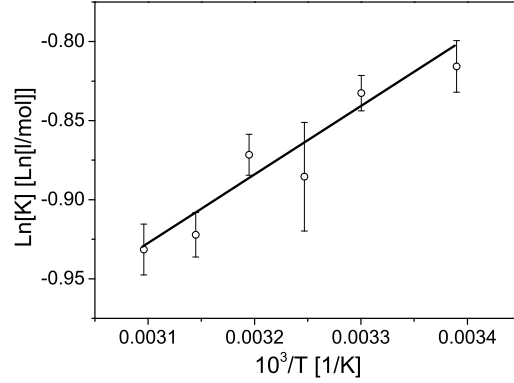


FIG. S19: Dependence of $\ln K$ on $1/T$ for model 2s.

D. Model 3

In this model, it is assumed that there is only one bond per oxygen and one bond per NH_2 , there are no cyclic dimers and there is only one association constant. The range of allowed aggregates with aggregation numbers up to $i = 3$ is shown in Figure S20; aggregates are chain-like in this case.

The number of ways to form bonds in model 3 can be written as

$$\Xi = \left(\frac{N!}{(N-M)!} \right)^2 \frac{4^M}{M!}, \quad (\text{S46})$$

and minimization of the free energy gives

$$\frac{M}{4(N-M)^2} = \frac{K}{V}. \quad (\text{S47})$$

In this case, the concentration of aggregates of size i is

$$c_i = 4^{i-1} K^{i-1} c_1^i, \quad (\text{S48})$$

so for the total concentration we have

$$c = \frac{c_1}{(1 - 4Kc_1)^2}. \quad (\text{S49})$$

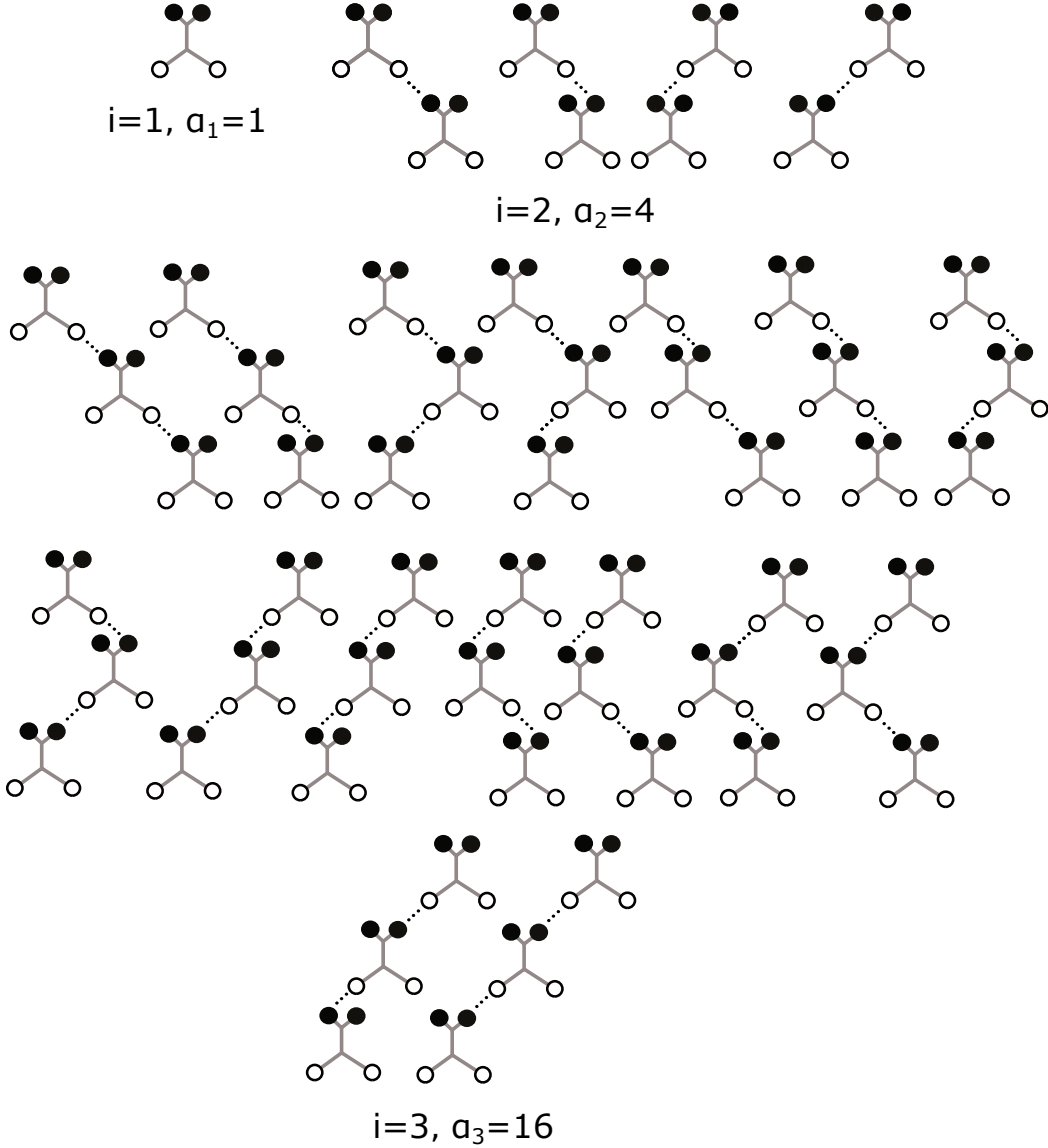


FIG. S20: Schematic representation of allowed aggregates with size up to $i = 3$ in model 3.

The concentration of free groups is given by $n_f = n - m$, so the relation between the total concentration and the concentration of free groups is

$$n = n_f + 4Kn_f^2. \quad (\text{S50})$$

1. Model 3m

Let us first assume that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in unimers. In this case, $c_1 = Ax$, where x is the height of the peak and A is some constant. The fitting equation is then

$$c = \frac{Ax}{(1 - 4KAx)^2}. \quad (\text{S51})$$

This expression is the same as the fitting equation for model 2g, but with an association constant that is two times smaller. This means that the fitting results are the same, with the values of the model parameters and statistical measures being $\text{AICc} = -499.7$, $\ln C = -1.9 \pm 0.7 \ln[\text{l/mol}]$, $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$, and $r^2 = 0.38$.

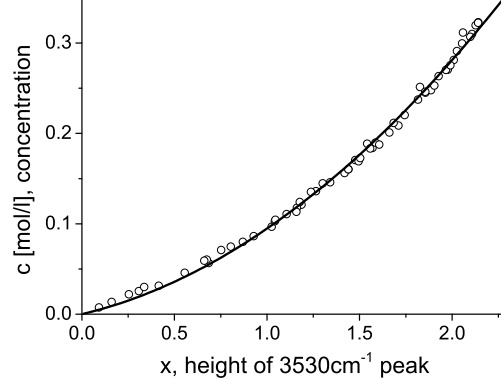


FIG. S21: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 3g.

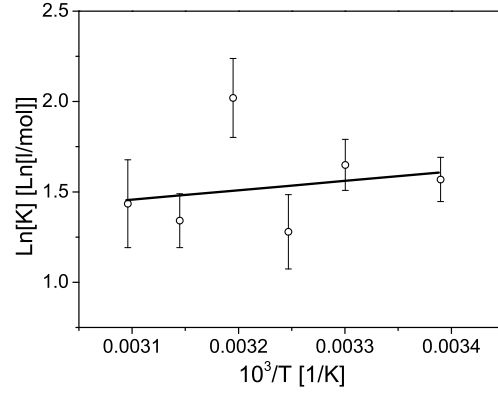


FIG. S22: Dependence of $\ln K$ on $1/T$ for model 3g.

2. Model 3g

In this model, it is assumed that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the free NH_2 groups. Here,

$$c = Ax + 4KA^2x^2. \quad (\text{S52})$$

The results of the fitting procedure are shown in Figures S21 and S22. The value of the AICc parameter for the fit in Figure S21 is -457.6 , and the estimates of the model parameters given by the fit in S22 are $\ln C = -0.15 \pm 3 \ln[\text{l/mol}]$ and $\epsilon = -1.0 \pm 1.9 \text{ kcal/mol}$, with $r^2 = 0.07$.

3. Model 3s

Let us first assume that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of free NH_2 groups in unimers and dimers. In this case, $Ax = c_1 + 4Kc_1^2$, where x is the height of the peak and A is some constant. The fitting equation in this case is

$$c = \frac{-1 + \sqrt{1 + 16KAx}}{2K(-3 + \sqrt{1 + 16KAx})^2}. \quad (\text{S53})$$

The results of the fitting procedure are shown in Figures S23 and S24, and the values of the model parameters and statistical measures are $\text{AICc} = -499.3$, $\ln C = 1.58 \pm 0.8 \ln[\text{l/mol}]$, $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$, and $r^2 = 0.37$.

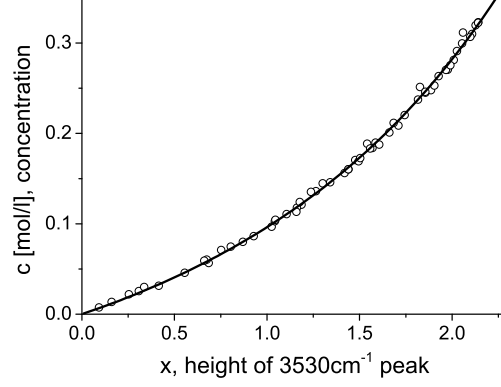


FIG. S23: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 3s.

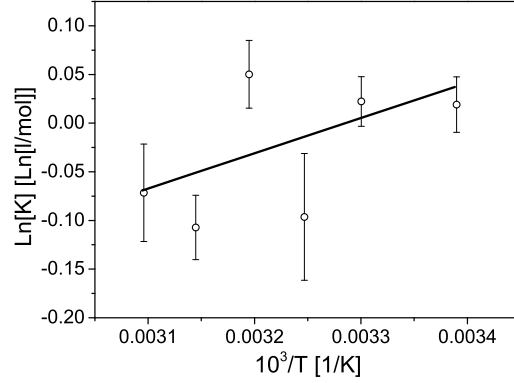


FIG. S24: Dependence of $\ln K$ on $1/T$ for model 3s.

The free energy density of hydrogen bonding in model 3, written in terms of the volume fraction of hydrogen bonding molecules $\phi = cv$ and the dimensionless association constant $K' = K/v$, has the form

$$f_{\text{HB}} = m + 2\phi \ln \frac{(\phi - m)}{\phi}, \quad (\text{S54})$$

where m is a solution of the equation $m/4(\phi - m)^2 = K'$.

E. Model 4

Model 4 is analogous to model 1 but with one bond allowed per NH_2 group and two bonds allowed per O in each acrylamide molecule (see Figure S25). This means that the expression for the relation between c and c_1 is the same in model 4 as in model 1. However, the number of free NH_2 groups in model 4 is different from model 1 and is equal to the number of aggregates as there is only one free NH_2 per aggregate. We can then write that

$$n_f = \sum_{i=1}^{\infty} c_i = \sum_{i=1}^{\infty} \frac{2^{i-1} (2i)!}{i! (i+1)!} K^{i-1} c_1^i = \frac{1 - 4c_1 K - \sqrt{1 - 8c_1 K}}{8c_1 K^2}. \quad (\text{S55})$$

Substituting this expression into Equation S19 gives

$$c = \frac{n_f + 2K n_f^2}{1 - 2K n_f}. \quad (\text{S56})$$

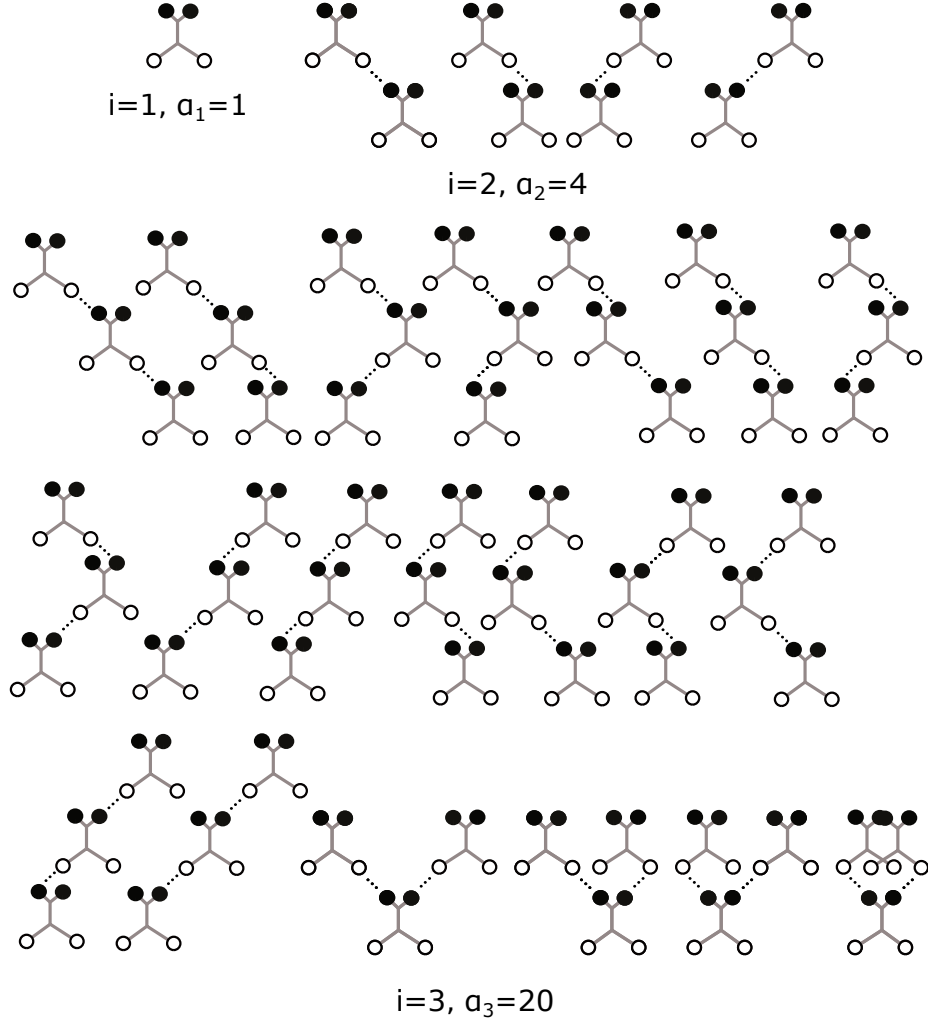


FIG. S25: Schematic representation of possible aggregates with size up to $i = 3$ in model 4.

1. Model 4m

First, we assume that the 3530cm^{-1} peak corresponds to out-of-phase vibrations of free acrylamide molecules. The fitting equation is the same as in model 1m, so we do not repeat the fitting procedure here.

2. Model 4g

With the free groups assumption, we can write fitting equation as

$$c = \frac{Ax + 2KA^2x^2}{1 - 2KAx}. \quad (\text{S57})$$

The results of the fit are shown in Figures S26 and S27. The quality of the fit in Figure S26 is characterized by the parameter $\text{AICc} = -495.8$, and the estimates for the model parameters resulting from the fit in Figure S27 are $\ln C = -0.8 \pm 0.9 \ln[\text{l/mol}]$ and $\epsilon = -0.7 \pm 0.5 \text{ kcal/mol}$, with $r^2 = 0.28$.

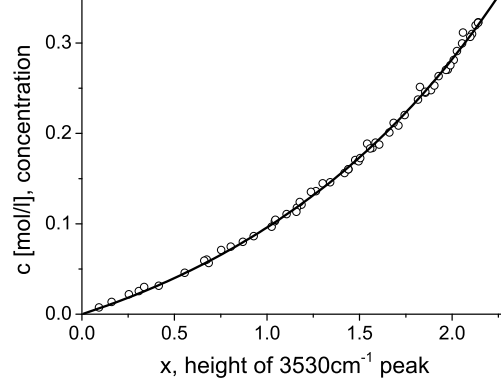


FIG. S26: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 4g.

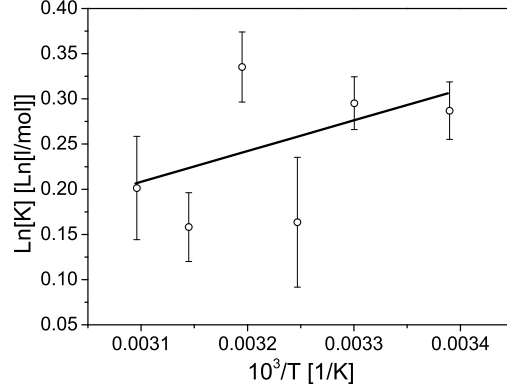


FIG. S27: Dependence of $\ln K$ on $1/T$ for model 4g.

3. Model 4s

In this case, we assume that the 3530cm^{-1} peak corresponds to out-of-phase vibrations of free NH_2 groups in unimers and dimers. The fitting equation and all fitting results are the same as for model 1s.

F. Model 5

Model 5 is the first model with "cooperativity" we consider. In this model, we allow one bond per oxygen, two bonds per NH_2 group, no cycles and two association constants (corresponding to bond energies ϵ_1 and ϵ_2) that depend on the bonding state of the NH_2 group in the donor molecule (see Figure S28).

Let us denote the number of molecules with donors involved in ϵ_1 bonds as N_1 (or, in other words, the number of donor molecules with one bonded hydrogen) and the number of molecules involved in ϵ_2 bonds as N_2 (in other words, the number of donor molecules with two bonded hydrogens). Correspondingly, the number of ϵ_1 bonds is $M_1 = N_1$ and number of ϵ_2 bonds is $M_2 = 2N_2$.

The number of ways to form M_1 bonds with energy ϵ_1 and $2N_2$ bonds with energy ϵ_2 can be written as

$$\Xi = \frac{N!2^{N_1+2N_2}}{(N - N_1 - 2N_2)! (N - N_1 - N_2)!N_1!N_2!}, \quad (\text{S58})$$

where the first factor is the number of ways to choose an acceptor for $M = N_1 + 2N_2$ bonds, and the second factor is the number of ways to choose N_1 and N_2 donor groups out of N molecules, taking into account the fact that in molecules with only one bonded hydrogen, this hydrogen can be chosen in two ways.

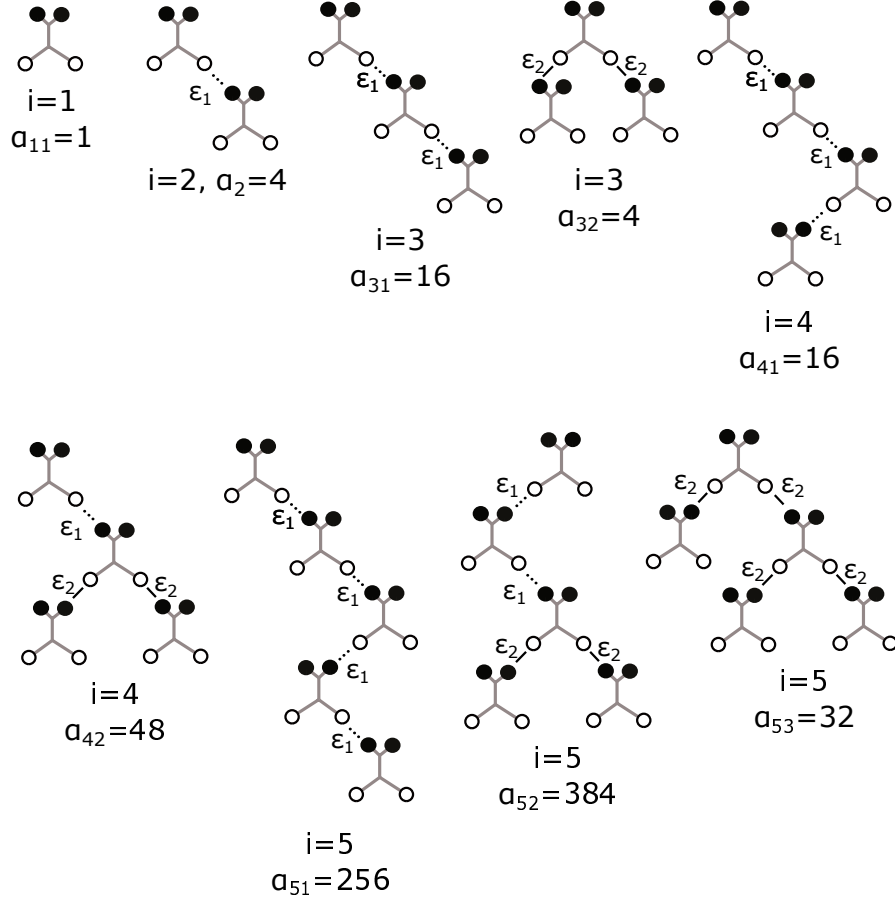


FIG. S28: Schematic representation of aggregates with size up to $i = 5$ in model 5. In contrast to the figures for one-parameter models, only configurations of aggregates with different energies are shown. The quantity a_{ij} gives the number of different configurations for an aggregate of i molecules that contains $2(j-1)$ ϵ_2 bonds.

The free energy of hydrogen bonding in model 5 is

$$F_{\text{HB}} = \epsilon_1 M_1 + \epsilon_2 M_2 - kT \ln \left(p_1^{M_1} p_2^{M_2} \Xi \right), \quad (\text{S59})$$

and minimizing this with respect to M_1 and M_2 gives

$$\frac{N_1}{4(N - N_1 - 2N_2)(N - N_1 - N_2)} = \frac{K_1}{V} \quad (\text{S60})$$

and

$$\frac{N_2}{4(N - N_1 - N_2)(N - N_1 - 2N_2)^2} = \frac{K_2^2}{V^2}. \quad (\text{S61})$$

It is interesting to note that the following equality exists:

$$\frac{(2N)!}{(2N - M)!M!} = \sum_{N_2=0}^{M/2} \frac{N!2^{M-2N_2}}{(N - M + N_2)!(M - 2N_2)!N_2!}, \quad (\text{S62})$$

which shows that model 5 reduces to model 1 when $\epsilon_1 = \epsilon_2$ – a property that we make use of in our calculations on model 1g above.

Now, we look for a relation between the total concentration and the concentration of unimers, and assume that concentration of aggregates of size i with $2(j-1)$ ϵ_2 -bonds has the form

$$c_{ij} = \alpha_{ij} K_2^{2(j-1)} K_1^{i-1-2(j-1)} c_1^i, 1 \leq j \leq (i+1)/2. \quad (\text{S63})$$

Substituting this expression in Equations S60 and S61 gives

$$\alpha_{ij} = \frac{2^{2i-2j}}{(j-1)!j!} \frac{(i-1)!}{(i-2j+1)!}, \quad (\text{S64})$$

and the dependence of the total solution concentration on the concentration of unimers can then be calculated to be

$$c = \sum_{i=1}^{\infty} \sum_{j=1}^{(i+1)/2} i \alpha_{ij} K_2^{2(j-1)} K_1^{i-1-2(j-1)} c_1^i = \frac{1 - \sqrt{1 - \frac{16K_2^2 c_1^2}{(1-4K_1 c_1)^2}}}{8K_2^2 c_1 \sqrt{1 - \frac{16K_2^2 c_1^2}{(1-4K_1 c_1)^2}}}. \quad (\text{S65})$$

Alternatively, the coefficients α_{ij} can be found from the generating function for the family of trees shown in Figure 4. If we denote $c_1 K_1$ as z_1 and $c_1 K_2$ as z_2 , we can write down the equation for the generating function as

$$G = 1 + 4Gz_1 + 4z_2^2 G^2, \quad (\text{S66})$$

which can be solved to find

$$G = \frac{1 - 4z_1 \pm \sqrt{(1 - 4z_1)^2 - 16z_2^2}}{8z_2^2} \quad (\text{S67})$$

where the required expression is that with the negative root. It is straightforward to verify that expansion of this expression in powers of z_1 and z_2 will yield the values of α_{ij} given by Equation S64. We also can see that if we put $z_1 = z_2$ we will recover the generating function for model 1, as would be expected from the fact that the two models are equivalent when $\epsilon_1 = \epsilon_2$.

The concentration of free groups is given by $n_f = c - n_1 - n_2$ and eliminating n_1 and n_2 from Equations S60 and S61 gives

$$c = \frac{1 - 4K_1 n_f + 16K_2^2 n_f^2 - (1 - 4K_1 n_f) \sqrt{1 + 16K_2^2 n_f^2}}{8K_2^2 n_f}. \quad (\text{S68})$$

1. Model 5m

Let us assume first that the 3530cm^{-1} peak corresponds to out-of-phase vibrations of the NH_2 group in free molecules. In this case, the fitting equation is

$$c = \frac{1 - \sqrt{1 - \frac{16A^2 K_2^2 x^2}{(1-4AK_1 x)^2}}}{8AK_2^2 x \sqrt{1 - \frac{16A^2 K_2^2 x^2}{(1-4AK_1 x)^2}}}. \quad (\text{S69})$$

The results of the fit are shown in Figures S29 and S30. The value of the AICc parameter for the nonlinear fit in Figure S29 is -501.7 , and the estimates of the model parameters given by the fits in Figure S30 are $\ln C_1 = -19.6 \pm 5 \ln[\text{l/mol}]$, $\epsilon_1 = -11 \pm 3 \text{ kcal/mol}$, $\ln C_2 = 3.4 \pm 1 \ln[\text{l/mol}]$, and $\epsilon_2 = 2.4 \pm 0.7 \text{ kcal/mol}$. In all our two-parameter models, we have two coefficients of determination, which in this case are given by $r_1^2 = 0.76$ and $r_2^2 = 0.73$.

2. Model 5g

Now we assume that the 3530cm^{-1} peak corresponds to out-of-phase vibrations of free groups NH_2 . In this case the fitting equation is

$$c = \frac{1 - 4AK_1 x + 16A^2 K_2^2 x^2 - (1 - 4AK_1 x) \sqrt{1 + 16A^2 K_2^2 x^2}}{8AK_2^2 x}. \quad (\text{S70})$$

In this case, we were unable to obtain any results, as the fitting procedure did not converge.

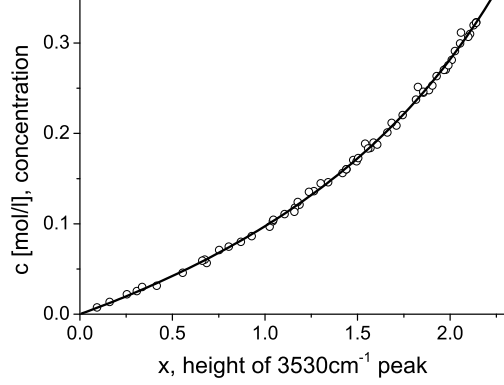


FIG. S29: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 5m.

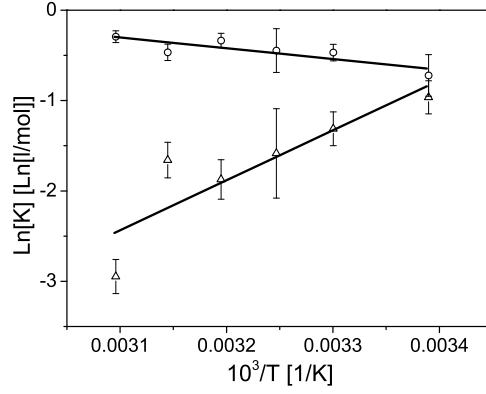


FIG. S30: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 5m.

3. Model 5s

We also check the possibility that the 3530cm^{-1} peak does not correspond to a single species (such as all free molecules or all free groups), but instead corresponds to absorption by free groups in some subset of aggregates. Here we check the subset composed of unimers and dimers, so $Ax = c_1 + 4K_1c_1^2$.

The results of the fit are shown in Figures S31 and S32. The value of AICc for the fit in Figure S31 is -501.7 , and the estimates of the model parameters given by the fits in S32 are $\ln C_1 = -21.5 \pm 6 \ln[\text{l/mol}]$, $\epsilon_1 = -12.5 \pm 3.5 \text{ kcal/mol}$, $\ln C_2 = 1.95 \pm 1 \ln[\text{l/mol}]$, and $\epsilon_2 = 1.3 \pm 0.6 \text{ kcal/mol}$, with $r_1^2 = 0.76$ and $r_2^2 = 0.56$.

G. Model 6

In model 6, we allow one bond per oxygen, two bonds per NH_2 group, no cycles and two association constants. However, in contrast to model 5, the association constant is now determined by the bonding state of the acceptor in the donor molecule (see Figure S33). Let the bond energy be denoted by ϵ_1 in the case when the oxygen in the donor amide group is free and by ϵ_2 otherwise. Then, the number of bonds with energy ϵ_1 is M_1 , the number of bonds with energy ϵ_2 is M_2 , the number of molecules with donors involved in ϵ_1 bonds is N_1 , and the number of molecules involved in ϵ_2 bonds is N_2 . The number of ways to form bonds can then be written as

$$\Xi = \frac{N! 2^{M_1+M_2}}{(N - M_1 - M_2)!} \frac{(2N - 2M_1 - 2M_2)! (2M_1 + 2M_2)!}{(2N - 3M_1 - 2M_2)! (2M_1 + M_2)!} \frac{1}{M_1! M_2!}, \quad (\text{S71})$$

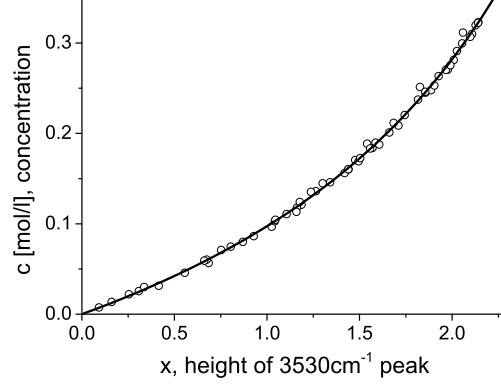


FIG. S31: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 5s.

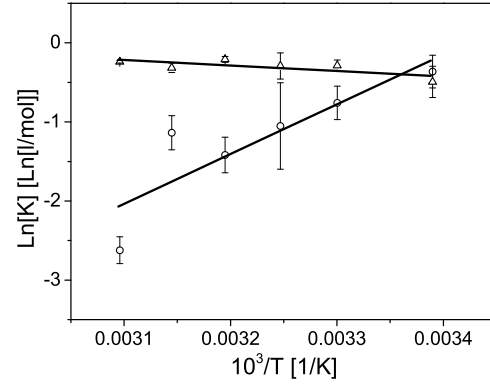


FIG. S32: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 5s.

where the first factor in Ξ is the number of ways to choose an acceptor for $M_1 + M_2$ bonds. This uses the assumption that there is only one bond per oxygen. The second factor is the number of ways to choose a donor for ϵ_1 bonds, with $2N - 2M_1 - 2M_2$ giving the number of hydrogens in molecules with free acceptor groups. The third factor is the number of ways to choose a donor for ϵ_2 bonds. The hydrogens for these bonds should be chosen from molecules with bonded acceptors. The number of such molecules is $M_1 + M_2$ and they contain $2M_1 + 2M_2$ hydrogens. As usual, the last term accounts for the indistinguishability of the bonds.

Minimization of the free energy gives

$$\frac{M_1 (2M_1 + M_2)^2 (N - M_1 - M_2)}{2 (M_1 + M_2)^2 (2N - 3M_1 - 2M_2)^3} = \frac{K_1}{V} \quad (\text{S72})$$

$$\frac{M_2 (2M_1 + M_2) (N - M_1 - M_2)}{2 (M_1 + M_2)^2 (2N - 3M_1 - 2M_2)^2} = \frac{K_2}{V}. \quad (\text{S73})$$

We can notice that there are two types of aggregates: those with one ϵ_1 bond and those with two ϵ_1 bonds. Then, the concentrations of aggregates with size i and either one or two ϵ_1 bonds can be written as

$$c_{i1} = \alpha_{i1} K_1 K_2^{i-2} c_1^i, \quad i \geq 2 \quad (\text{S74})$$

$$c_{i2} = \alpha_{i2} K_1^2 K_2^{i-3} c_1^i, \quad i \geq 3, \quad (\text{S75})$$

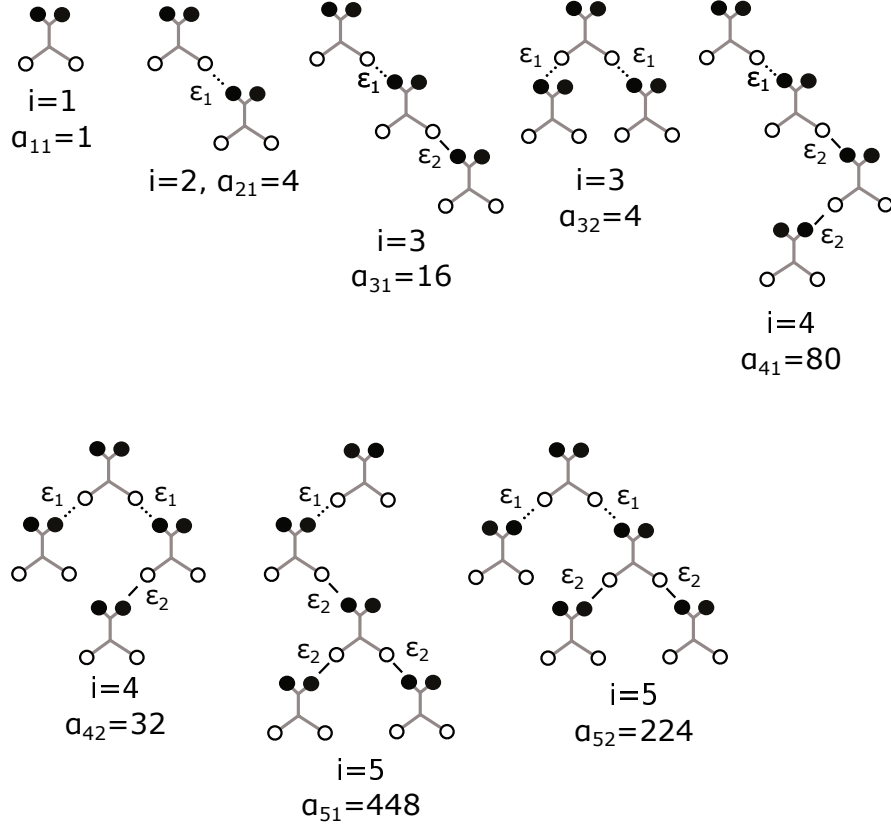


FIG. S33: Schematic representation of aggregates with size up to $i = 5$ in model 6. Only aggregates with different total bond energies are shown.

and the total concentration of acrylamide and concentrations of each type of bond as

$$c = c_1 + \sum_{i=2}^{\infty} i c_{i1} + \sum_{i=3}^{\infty} i c_{i2} \quad (\text{S76})$$

$$m_1 = \sum_{i=2}^{\infty} c_{i1} + \sum_{i=3}^{\infty} 2 c_{i2} \quad (\text{S77})$$

$$m_2 = \sum_{i=2}^{\infty} (i-2) c_{i1} + \sum_{i=3}^{\infty} (i-3) c_{i2}. \quad (\text{S78})$$

Substituting these expressions into Equations S72 and S73 leads to the following expressions for $\alpha_{i1,2}$:

$$\alpha_{i1} = 2^i \frac{(2i-2)!}{(i-1)!i!}, i \geq 2 \quad (\text{S79})$$

$$\alpha_{i2} = 2^{i+1} \frac{(2i-3)!}{(i-3)!(i+1)!}, i \geq 3. \quad (\text{S80})$$

For the dependence of the total concentration on the concentration of unimers we have

$$c = \frac{16c_1^2 K_1 K_2^3 - 16c_1^2 K_2^2 K_1^2 + K_1^2 - 4c_1 K_1^2 K_2}{8c_1 K_2^4 \sqrt{1 - 8c_1 K_2}} + \frac{8c_1^2 K_2^4 - 16c_1^2 K_1 K_2^3 - K_1^2 + 8c_1^2 K_1^2 K_2^2}{8c_1 K_2^4}. \quad (\text{S81})$$

Another way to determine the coefficients α_{i1} and α_{i2} is to find an expression for the generating function, as was done in for models 1 and 5. In the current case, we have

$$G(z_1, z_2) = 1 + 4z_1 G_1(z_2) + 4z_1^2 G_1^2(z_2), \quad (\text{S82})$$

where

$$G_1(z_2) = \frac{1 - 4z_2 - \sqrt{1 - 8z_2}}{8z_2^2}, \quad (\text{S83})$$

the generating function for model 1.

Now, let us turn to the calculation of the number of free groups. In order to do this, it is necessary to distinguish molecules with one hydrogen bond per NH_2 group and two hydrogen bonds per NH_2 group. We denote the number of molecules with one bonded hydrogen and free oxygen by N_1 (Figure S34) and the number of molecules with two bonded hydrogens and free oxygen by N_3 , so that $M_1 = N_1 + 2N_3$. Similarly, we denote the number of molecules with one bonded hydrogen and bonded oxygen by N_2 , and the number of molecules with two bonded hydrogens and bonded oxygen, by N_4 , so that $M_2 = N_2 + 2N_4$. Then, the number of ways to form bonds is

$$\Xi = \frac{N! 2^{M_1+M_2}}{(N - M_1 - M_2)!} \frac{(N - M_1 - M_2)! 2^{N_1}}{(N - M_1 - M_2 - N_1 - N_3)! N_1! N_3!} \frac{(M_1 + M_2)! 2^{N_2}}{(M_1 + M_2 - N_2 - N_4)! N_2! N_4!}. \quad (\text{S84})$$

After simplification, this becomes

$$\Xi = \frac{N! (N_1 + N_2 + 2N_3 + 2N_4)! 4^{N_1+N_2+N_3+N_4}}{(N - 2N_1 - 3N_3 - N_2 - 2N_4)! (N_1 + 2N_3 + N_4)! N_1! N_2! N_3! N_4!}. \quad (\text{S85})$$

Also, for simplicity, we assume first that there are four association constants and correspondingly four different bond energies, so we can write for the free energy of hydrogen bonding

$$F_{\text{HB}} = \epsilon_1 N_1 + \epsilon_2 N_2 + 2\epsilon_3 N_3 + 2\epsilon_4 N_4 - kT \ln \left(p_1^{N_1} p_2^{N_2} p_3^{2N_3} p_4^{2N_4} \Xi \right). \quad (\text{S86})$$

Minimization of this expression gives

$$\frac{N_1 (N_1 + 2N_3 + N_4)}{4 (N_1 + N_2 + 2N_3 + 2N_4) (N - 2N_1 - 3N_3 - N_2 - 2N_4)^2} = \frac{K_1}{V} \quad (\text{S87})$$

$$\frac{N_2}{4 (N_1 + N_2 + 2N_3 + 2N_4) (N - 2N_1 - 3N_3 - N_2 - 2N_4)} = \frac{K_2}{V} \quad (\text{S88})$$

$$\frac{N_3 (N_1 + 2N_3 + N_4)^2}{4 (N_1 + N_2 + 2N_3 + 2N_4)^2 (N - 2N_1 - 3N_3 - N_2 - 2N_4)^3} = \frac{K_3^2}{V^2} \quad (\text{S89})$$

$$\frac{N_4 (N_1 + 2N_3 + N_4)}{4 (N_1 + N_2 + 2N_3 + 2N_4)^2 (N - 2N_1 - 3N_3 - N_2 - 2N_4)^2} = \frac{K_4^2}{V^2}, \quad (\text{S90})$$

and the number of free groups is given by $N_f = N - N_1 - N_2 - N_3 - N_4$.

We notice that, since only one bond per oxygen is allowed, all aggregates have either one ϵ_1 bond or two ϵ_3 bonds, which lie at the "root" of each aggregate. Furthermore, we note that the distribution of ϵ_2 and ϵ_4 bonds is very similar to the original model 6. Then, we can label different aggregates by the set of three numbers $\{1ij\}$ or $\{3ij\}$ where the first letter denotes the type of "root" (type 1 bonds or type 3 bonds), i is the number of molecules in the aggregate and j the number of type 4 bonds in aggregate. The concentrations of aggregates can then be written as

$$c_{1ij} = \alpha_{1ij} K_1 K_2^{i-2j} K_4^{2j-2} c_1^i \quad (\text{S91})$$

$$c_{3ij} = \alpha_{3ij} K_3^2 K_2^{i-1-2j} K_4^{2j-2} c_1^i \quad (\text{S92})$$

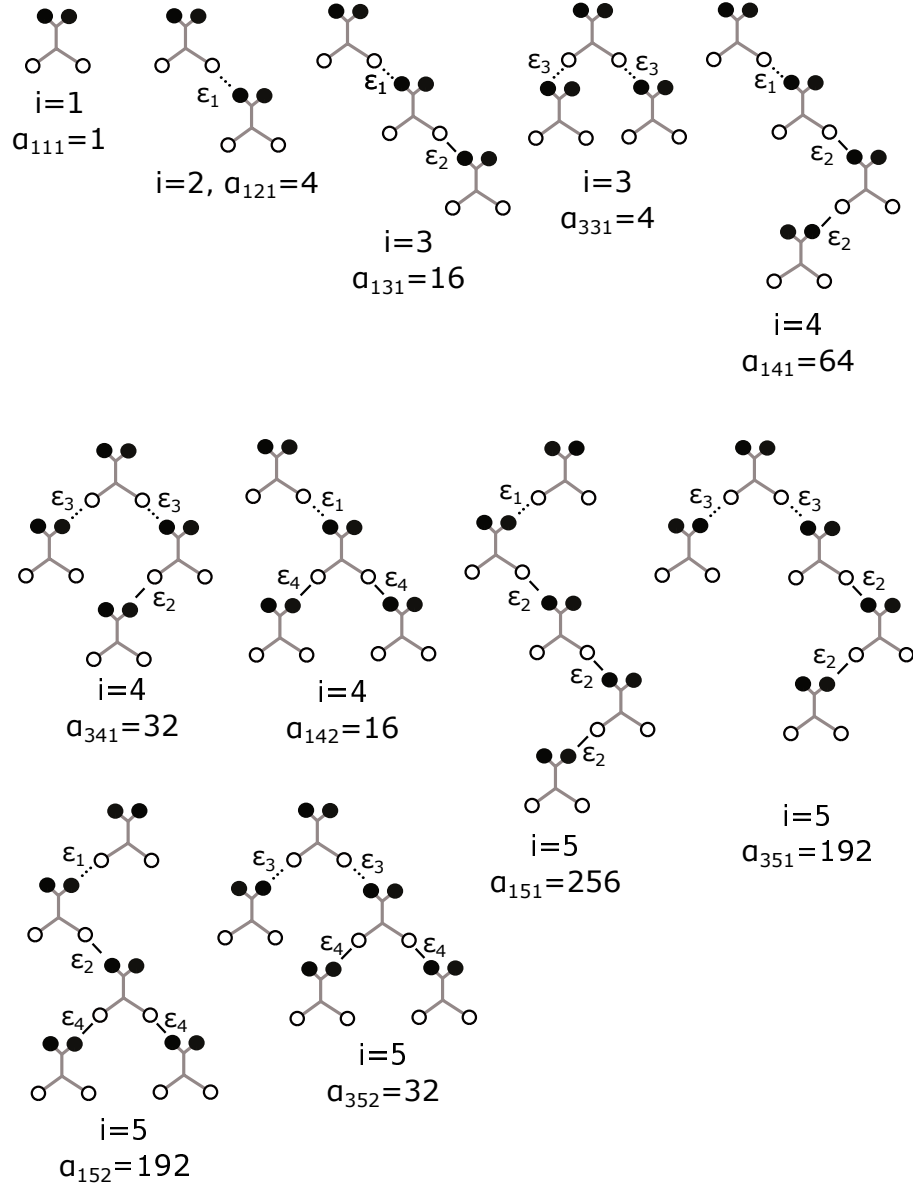


FIG. S34: Schematic representation of aggregates with size up to $i = 5$ in the modification of model 6 with four association constants. Only aggregates with different total bond energies are shown.

and the concentrations of molecules in different bonding states can be calculated to be

$$c = c_1 + \sum_{i=2}^{\infty} \sum_{j=1}^{i/2} i c_{1ij} + \sum_{i=3}^{\infty} \sum_{j=1}^{(i-1)/2} i c_{3ij} \quad (\text{S93})$$

$$c_1 = \sum_{i=2}^{\infty} \sum_{j=1}^{i/2} c_{1ij} = \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} c_{1ij} \quad (\text{S94})$$

$$c_3 = \sum_{i=3}^{\infty} \sum_{j=1}^{(i-1)/2} c_{3ij} = \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} c_{3ij} \quad (\text{S95})$$

$$c_2 = \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} (i-2j) c_{1ij} + \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} (i-1-2j) c_{3nj} \quad (\text{S96})$$

$$c_4 = \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} (j-1) c_{1ij} + \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} (j-1) C_{3ij}. \quad (\text{S97})$$

Additionally, we can write that

$$n_f = c_1 + \sum_{j=1}^{\infty} \sum_{i=2j}^{\infty} j c_{1ij} + \sum_{j=1}^{\infty} \sum_{i=2j+1}^{\infty} (j+1) c_{3ij}. \quad (\text{S98})$$

Substituting these expressions into Equations S87, S88, S89, S90 gives

$$\alpha_{1ij} = 2^{2i-2j} \frac{1}{j!(j-1)!} \frac{(i-2)!}{(i-2j)!}, i \geq 2j \quad (\text{S99})$$

$$\alpha_{3ij} = 2^{2i-2j-1} \frac{1}{(j+1)!(j-1)!} \frac{(i-2)!}{(i-2j-1)!}, i \geq 2j+1. \quad (\text{S100})$$

We can now evaluate the sums and calculate the dependence of n_f on c_1 .

Since we assumed that equilibrium is described by two association constants in model 6, we put $K_3 = K_1$ and $K_4 = K_2$ and find

$$n_f = \frac{c_1 (4c_1 K_1 K_2^2 + K_2^2 \sqrt{1-8c_1 K_2} + K_1^2 (1-4c_1 K_2 - \sqrt{1-8c_1 K_2}))}{K_2^2 \sqrt{1-8c_1 K_2}}. \quad (\text{S101})$$

As we know the dependence of both c and n_f on c_1 , we have parametrically defined a function $c(n_f)$, which we can use to fit experimental data in case of the free groups assumption.

1. Model 6m

Here, it is assumed that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in free molecules. In this case, we put $c_1 = Ax$ and substitute it into Equation S81.

The fitting results are shown in Figures S35 and S36. The value of the information criterion for the fit shown in Figure S35 is $\text{AICc} = -501.7$, and the estimates of the model parameters given by the fits in Figure S36 are $\ln C_1 = -9.9 \pm 1 \ln[\text{l/mol}]$, $\epsilon_1 = -5.2 \pm 0.8 \text{ kcal/mol}$, $\ln C_2 = -0.62 \pm 0.5 \ln[\text{l/mol}]$, and $\epsilon_2 = 0.1 \pm 0.2 \text{ kcal/mol}$, with $r_1^2 = 0.91$ and $r_2^2 = 0.03$.

2. Model 6g

In this section, it is assumed that the 3530cm^{-1} peak corresponds to absorption by free groups, so that $n_f = Ax$. The results of the fitting procedure are shown in Figures S37 and S38. For the fit in Figure S37, we have that $\text{AICc} = -493.8$. The estimates of the model parameters for the fits in Figure S38 are $\ln C_1 = -11.5 \pm 5 \ln[\text{l/mol}]$, $\epsilon_1 = -5.9 \pm 3 \text{ kcal/mol}$, $\ln C_2 = 3.5 \pm 0.6 \ln[\text{l/mol}]$, and $\epsilon_2 = -2.2 \pm 0.3 \text{ kcal/mol}$, with $r_1^2 = 0.71$ and $r_2^2 = 0.90$.

3. Model 6s

For the assumption $Ax = c_1 + 4K_1 c_1^2 + 4K_1^2 c_1^3$ (NH_2 groups in unimers, dimers and trimers without ϵ_2 -bonds), we have the fitting results shown in Figures S39 and S40. The quality of fit in Figure S39 can be characterized by $\text{AICc} = -501.7$. The estimates of parameters for the fits in Figure S40 are $\ln C_1 = -11.89 \pm 2 \ln[\text{l/mol}]$, $\epsilon_1 = -6.7 \pm 1.1 \text{ kcal/mol}$, $\ln C_2 = -3.18 \pm 0.3 \ln[\text{l/mol}]$, and $\epsilon_2 = -1.6 \pm 0.2 \text{ kcal/mol}$, with $r_1^2 = 0.91$ and $r_2^2 = 0.95$.

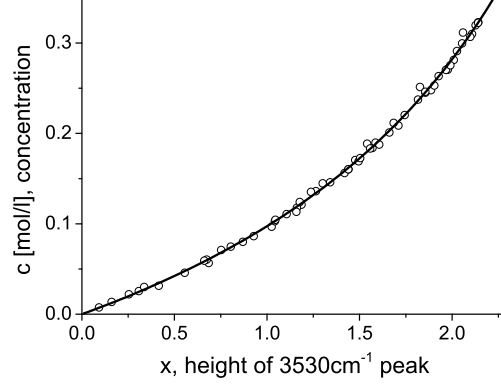


FIG. S35: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 6m.

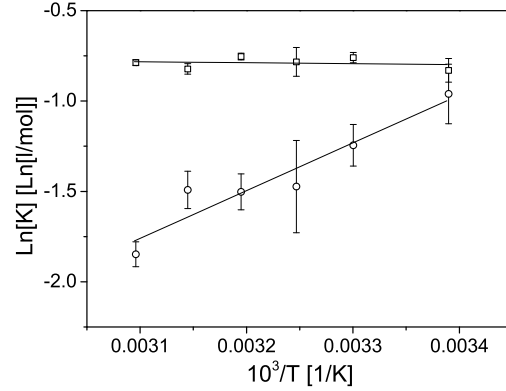


FIG. S36: Dependence of $\ln K_1$ (circles) and $\ln K_2$ (squares) on $1/T$ for model 6m.

The free-energy density of hydrogen bonding in model 6 in terms of the volume fraction of hydrogen bonding molecules $\phi = cv$ and dimensionless association constants $K'_1 = K_1/v$, $K'_2 = K_2/v$ has the form

$$f_{\text{HB}} = m_1 + m_2 + \phi \ln \frac{(2\phi - 3m_1 - 2m_2)^2}{4\phi(\phi - m_1 - m_2)}, \quad (\text{S102})$$

where m_1 and m_2 can be calculated by taking the sums in Equations S77 and S78 to be functions of the association constants and c_1 . Then, we can consider c_1 as a parameter and now have a parametrically defined function $f_{\text{HB}}(\phi)$.

H. Model 7

In model 7, we again allow one bond per oxygen, two bonds per NH_2 group, no cyclic dimers and two association constants. However, in contrast to models 5 and 6, the association constant is now determined by the bonding state of the NH_2 group in the acceptor molecule (see Figure S41).

We denote the energy of a bond formed by an acceptor with a free donor group by ϵ_1 and the energy of a bond formed by an acceptor molecule with one or two bonded hydrogens by ϵ_2 . As in some previous models, we will first consider a more detailed case (with three association constants), in which the situations corresponding to one bonded hydrogen in an acceptor molecule (ϵ_2) and to two bonded hydrogens in an acceptor molecule (ϵ_3) differ from each other (see Figure S42). We write M_1 for the number of bonds with energy ϵ_1 , M_2 for the number of bonds with energy ϵ_2 , and M_3 for the number of bonds with energy ϵ_3 (see Figure S42). Similarly, we write N_1 for the number of free molecules, N_2 for the number of molecules with a free oxygen and one bonded hydrogen, N_3 for the number

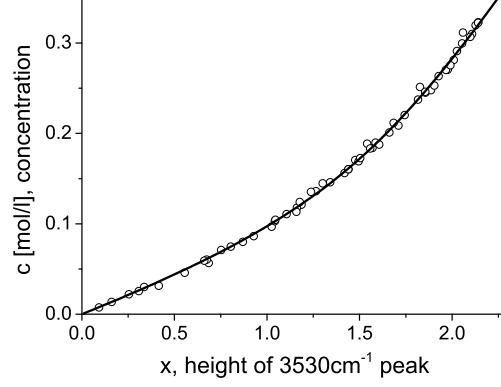


FIG. S37: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 6g.

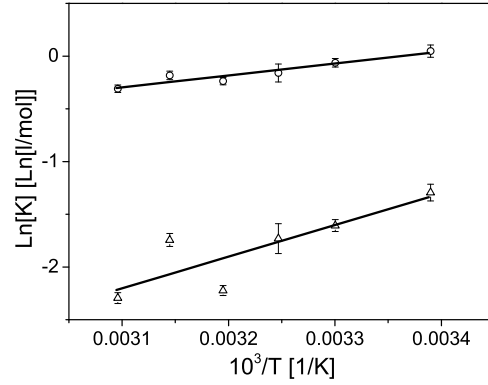


FIG. S38: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 6g.

of molecules with a free oxygen and two bonded hydrogens, N_4 for the number of molecules with a bonded oxygen and both hydrogens free, N_5 for the number of molecules with a bonded oxygen and one bonded hydrogen, and N_6 for the number of molecules with a bonded oxygen and two bonded hydrogens. Among these values, the following relations exist: $M_1 = N_4$, $M_2 = N_5$, $M_3 = N_6$, $N = \sum_{i=1}^6 N_i$, $M_1 + M_2 + M_3 = N_2 + 2N_3 + N_5 + 2N_6$.

In this case the number of ways to choose acceptors is

$$\frac{N!2^{N_4+N_5+N_6}}{(N - N_4 - N_5 - N_6)!} \quad (\text{S103})$$

The number of ways to choose donors is (this expression is effectively the same thing as Equation S62)

$$\frac{(N - N_4 - N_5 - N_6)!2^{N_2} (N_4 + N_5 + N_6)!2^{N_5}}{N_1!N_2!N_3! N_4!N_5!N_6!}, \quad (\text{S104})$$

and finally we have

$$\Xi = \frac{N!2^{2N-2N_1-2N_4} (N_4 + N_5 + N_6)!}{N_1!(2N - 2N_1 - 3N_4 - 2N_5 - N_6)!(N_1 + 2N_4 + N_5 - N)!N_4!N_5!N_6!}. \quad (\text{S105})$$

Minimization of the free energy yields the following set of equations:

$$\frac{4N_1 (N_1 + 2N_4 + N_5 - N)}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)^2} = 1 \quad (\text{S106})$$

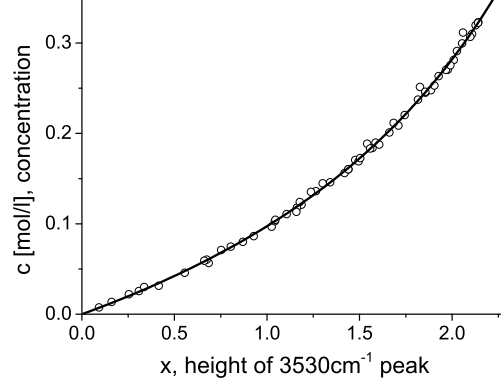


FIG. S39: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 6s.

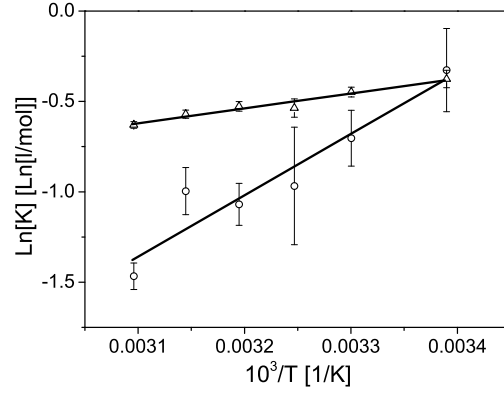


FIG. S40: Dependence of $\ln K_1$ (circles) and $\ln K_2$ (squares) on $1/T$ for model 6s.

$$\frac{4N_4(N_1 + 2N_4 + N_5 - N)^2}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)^3(N_4 + N_5 + N_6)} = \frac{K_1}{V} \quad (\text{S107})$$

$$\frac{N_5(N_1 + 2N_4 + N_5 - N)}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)^2(N_4 + N_5 + N_6)} = \frac{K_2}{V} \quad (\text{S108})$$

$$\frac{N_6}{(2N - 2N_1 - 3N_4 - 2N_5 - N_6)(N_4 + N_5 + N_6)} = \frac{K_3}{V}. \quad (\text{S109})$$

For the concentrations of aggregates, we may write

$$C_{ijk} = \alpha_{ijk} K_1^{i-1-j-k} K_2^j K_3^k c_1^i, \quad (\text{S110})$$

where $0 \leq j \leq i - 2$ and there are two series of k values that satisfy $2k + j + 2 = i$ for j and i both odd or even and $2k + j + 3 = i$ otherwise. Based on these relations, one index may be removed and two series of concentrations introduced instead:

$$C_{1ik} = \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i, \quad 0 \leq k \leq \infty, 2k + 2 \leq i \leq \infty \quad (\text{S111})$$

$$C_{2ik} = \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i, \quad 0 \leq k \leq \infty, 2k + 3 \leq i \leq \infty. \quad (\text{S112})$$

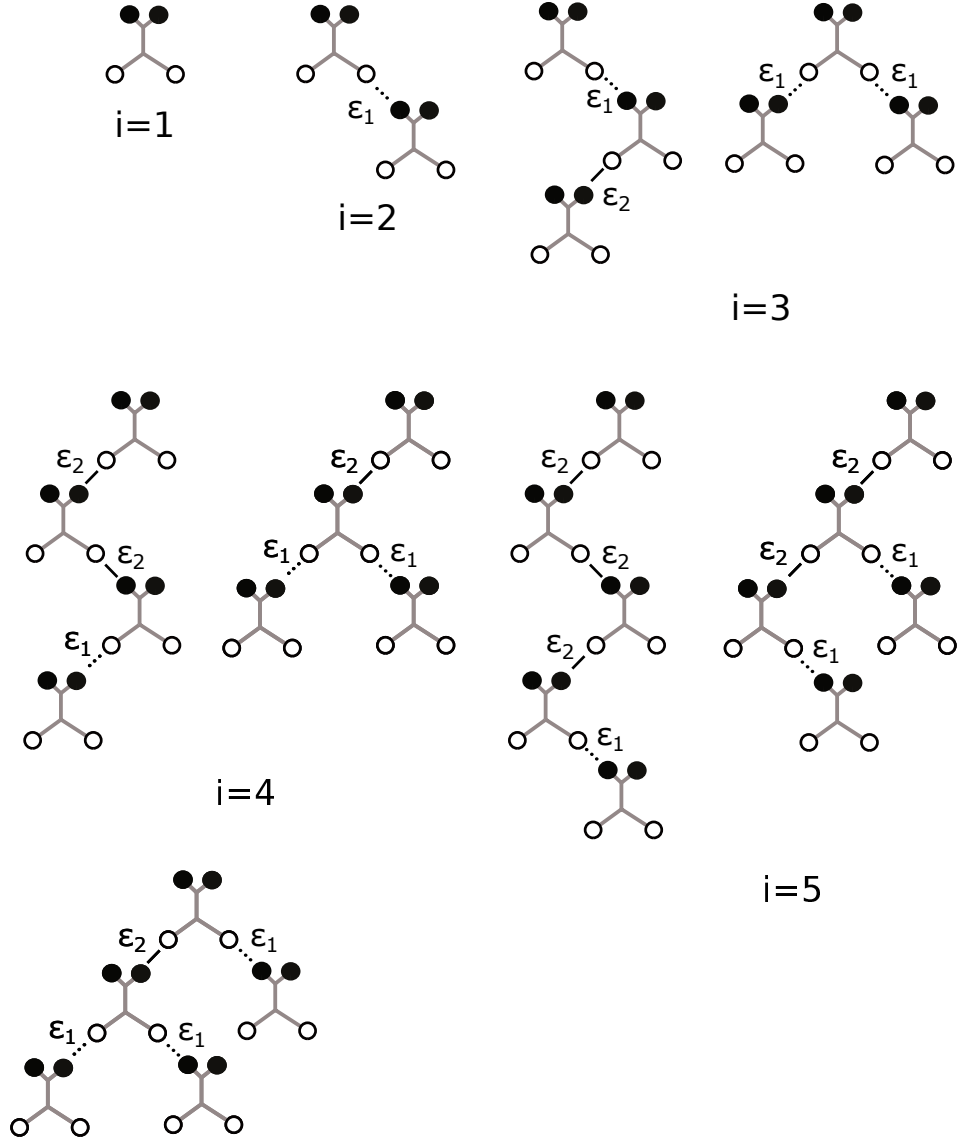


FIG. S41: Schematic representation of aggregates with size up to $i = 5$ in model 7. Only configurations corresponding to different energies are shown.

Therefore, we can write

$$c = c_1 + \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} i \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} i \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \quad (\text{S113})$$

$$n_1 = c_1 \quad (\text{S114})$$

$$n_4 = \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} (k+1) \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} (k+2) \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \quad (\text{S115})$$

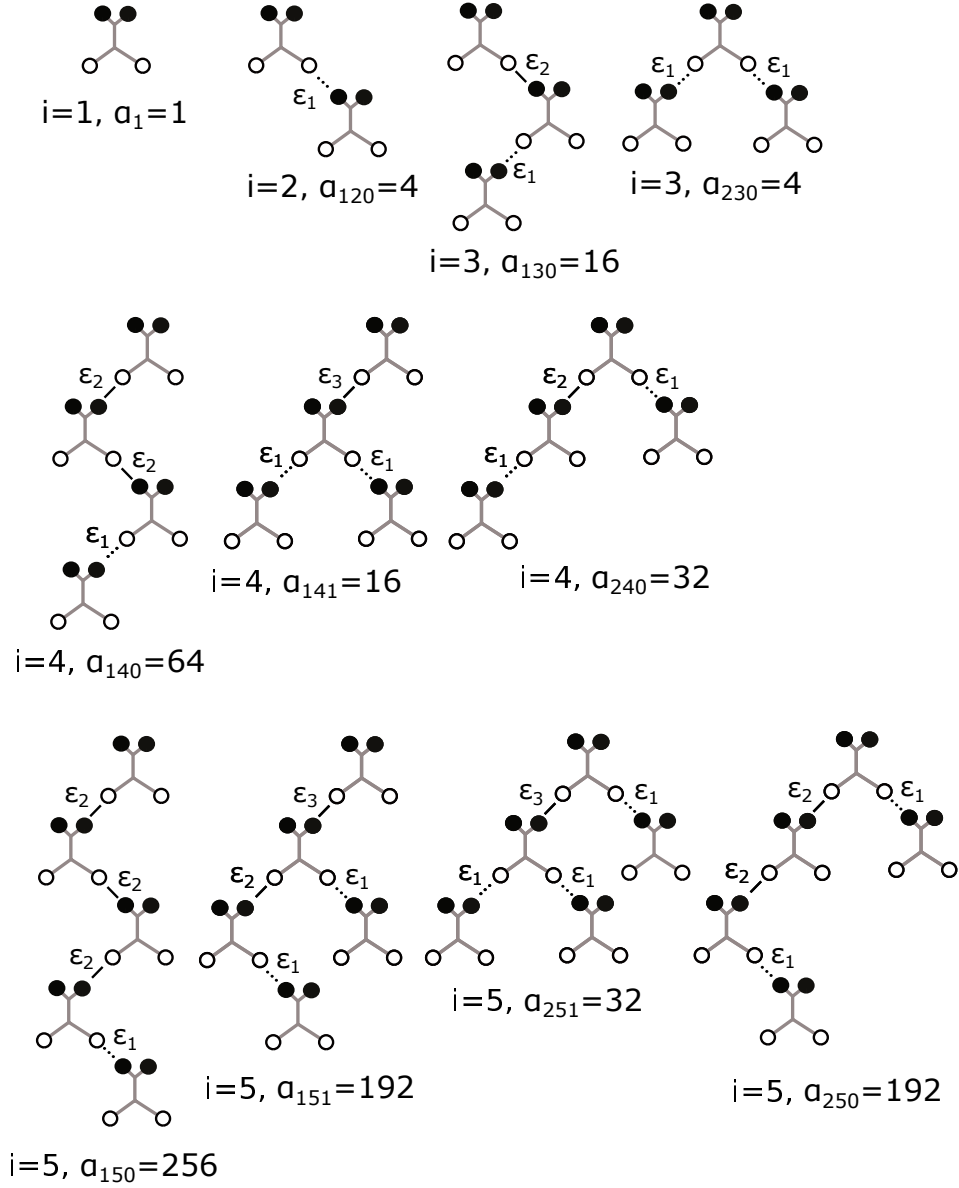


FIG. S42: Schematic representation of aggregates with size up to $i = 5$ in model 7. Only configurations corresponding to different energies are shown.

$$\begin{aligned}
 n_5 = & \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} (i-2k-2) \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i \\
 & + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} (i-2k-3) \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i
 \end{aligned} \tag{S116}$$

$$n_6 = \sum_{k=0}^{\infty} \sum_{i=2k+2}^{\infty} k \alpha_{1ik} K_1^{k+1} K_2^{i-2k-2} K_3^k c_1^i + \sum_{k=0}^{\infty} \sum_{i=2k+3}^{\infty} k \alpha_{2ik} K_1^{k+2} K_2^{i-2k-3} K_3^k c_1^i \tag{S117}$$

$$\alpha_{1ik} = 4^{i-1-k} \frac{(i-2)!}{(i-2-2k)! k! (k+1)!} \tag{S118}$$

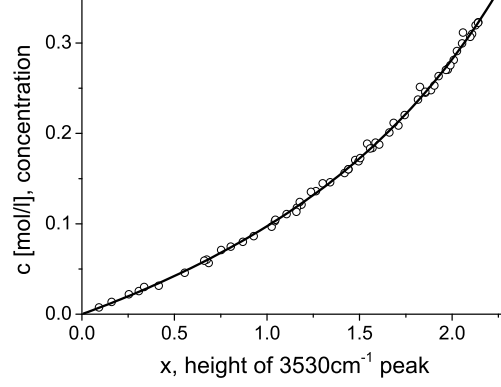


FIG. S43: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 7m.

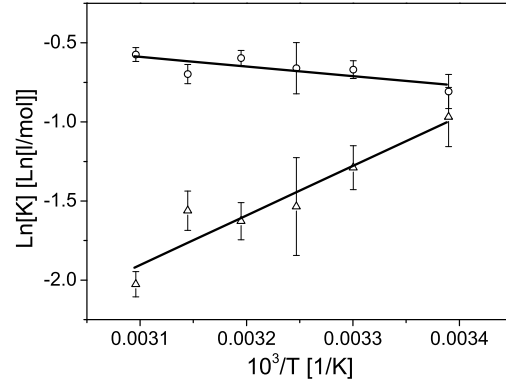


FIG. S44: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 7m.

$$\alpha_{2ik} = 4^{i-2-k} \frac{(i-2)!(2k+2)}{(i-3-2k)!(k+1)!(k+2)!} \quad (\text{S119})$$

By summing the series, assuming that $K_2 = K_3$, it can be shown that

$$c = \frac{1 - 4c_1K_2}{8c_1K_2^2\sqrt{1 - 8c_1K_2 - 16c_1^2(K_1 - K_2)K_2}} + \frac{8c_1^2K_2(K_2 - K_1) - 1}{8c_1K_2^2}. \quad (\text{S120})$$

Finally, the number of free groups, which is the sum of c_1 and n_4 , can be calculated as

$$n_f = c_1 - \frac{c_1K_1}{K_2} + \frac{c_1K_1}{K_2\sqrt{1 - 8c_1K_2 - 16c_1^2(K_1 - K_2)K_2}}. \quad (\text{S121})$$

1. Model 7m

Here, it is assumed that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in free molecules. In this case, we put $c_1 = Ax$ and substitute it into Equation S120.

The fitting results are shown in Figures S43 and S44. The quality of fit in Figure S43 can be characterized by $\text{AICc} = -501.7$, and the estimates of the model parameters for the fits in Figure S44 are $\ln C_1 = -11.6 \pm 2 \ln[\text{l/mol}]$, $\epsilon_1 = -6.2 \pm 1 \text{ kcal/mol}$, $\ln C_2 = 1.3 \pm 0.8 \ln[\text{l/mol}]$, and $\epsilon_2 = 1.2 \pm 0.5 \text{ kcal/mol}$, with $r_1^2 = 0.9$ and $r_2^2 = 0.62$.

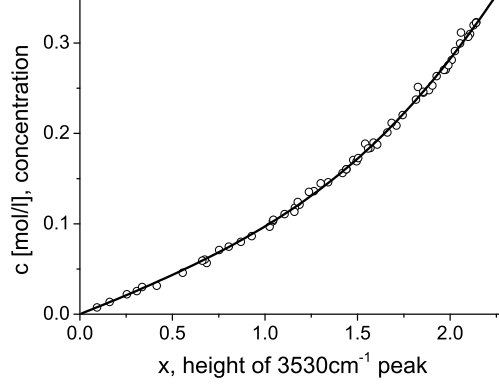


FIG. S45: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 7g.

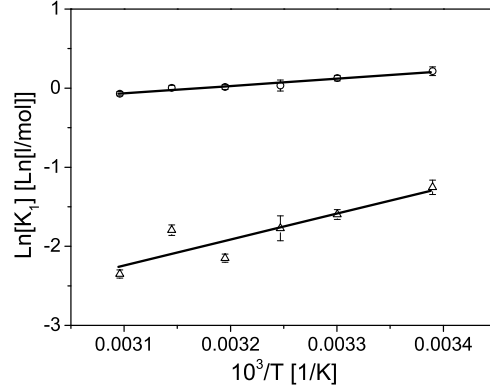


FIG. S46: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 7g.

2. Model 7g

Next, it is assumed that the 3530cm^{-1} peak corresponds to absorption by free groups, so that $n_f = Ax$.

The results of the fitting procedure are shown in Figures S45 and S46. The value of AICc for the fit in Figure S45 is -497.4 . The estimates of the model parameters for the fits in Figure S46 are $\ln C_1 = -12.4 \pm 3 \ln[\text{l/mol}]$, $\epsilon_1 = -6.5 \pm 1.6 \text{ kcal/mol}$, $\ln C_2 = -2.9 \pm 0.3 \ln[\text{l/mol}]$, and $\epsilon_2 = -1.8 \pm 0.2 \text{ kcal/mol}$, with $r_1^2 = 0.80$ and $r_2^2 = 0.95$.

3. Model 7s

In the case of the assumption $Ax = c_1 + 4K_1c_1^2 + 4K_1^2c_1^3$ (NH_2 groups in unimers, dimers and trimers without ϵ_2 -bonds) the fitting results shown in Figures S47 and S48 are found. The value of the information criterion for the fit in Figure S47 is $\text{AICc} = -501.7$, and the estimates of the model parameters given by the fits in Figure S48 are $\ln C_1 = -14.4 \pm 2 \ln[\text{l/mol}]$, $\epsilon_1 = -8.2 \pm 1.4 \text{ kcal/mol}$, $\ln C_2 = -1.5 \pm 0.3 \ln[\text{l/mol}]$, and $\epsilon_2 = -0.6 \pm 0.2 \text{ kcal/mol}$, with $r_1^2 = 0.9$ and $r_2^2 = 0.69$.

I. Model 8

In model 8, we allow only one bond per oxygen and one bond per NH_2 group, so all aggregates are supposed to be linear. We also assume that the association equilibrium is described by two association constants and there are no cyclic dimers. We suppose that the bond has energy ϵ_1 if the acceptor of donor molecule is free and ϵ_2 otherwise

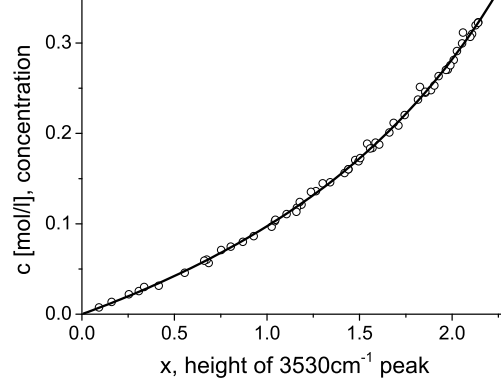


FIG. S47: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 7s.

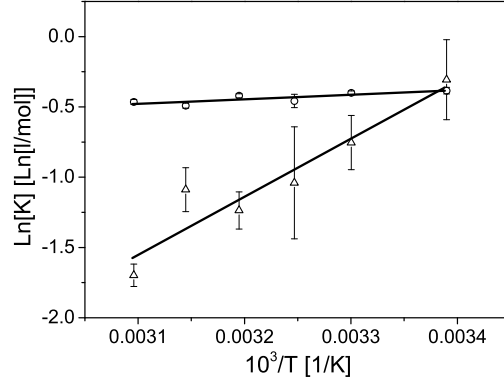


FIG. S48: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 7s.

(see Figure S49). It is worth mentioning that this model does not change if we assume that the energy of the bond is defined by the bonding state of the donor group in the acceptor molecule. The number of free molecules will be denoted by N_0 , the number of molecules with one bonded hydrogen and free oxygen by N_1 , the number of molecules with one bonded hydrogen and one bond per oxygen by N_2 , and the number of molecules with one bond per oxygen and free hydrogens by N_3 . Then, the number of ways to form bonds is

$$\Xi = \frac{N!2^{N_3+N_2}}{(N - N_3 - N_2)!} \frac{(N_0 + N_1)!2^{N_1}}{N_1!N_0!} \frac{(N_3 + N_2)!2^{N_2}}{N_2!N_3!}, \quad (\text{S122})$$

where the first factor is the number of ways to choose acceptors, the next factor is the number of ways to choose donors for bonds with energy ϵ_1 , and the following factor is the number of ways to choose bonds with energy ϵ_2 .

Taking into account that $N_3 = N_1$ and $N_0 = N - 2N_1 - N_2$, we have

$$\Xi = \frac{N!(N_1 + N_2)!2^{2N_1+2N_2}}{(N - 2N_1 - N_2)!(N_1!)^2 N_2!}. \quad (\text{S123})$$

Minimization of the free energy gives

$$\frac{N_1^2}{4(N_1 + N_2)(N - 2N_1 - N_2)^2} = \frac{K_1}{V} \quad (\text{S124})$$

$$\frac{N_2}{4(N_1 + N_2)(N - 2N_1 - N_2)} = \frac{K_2}{V}. \quad (\text{S125})$$

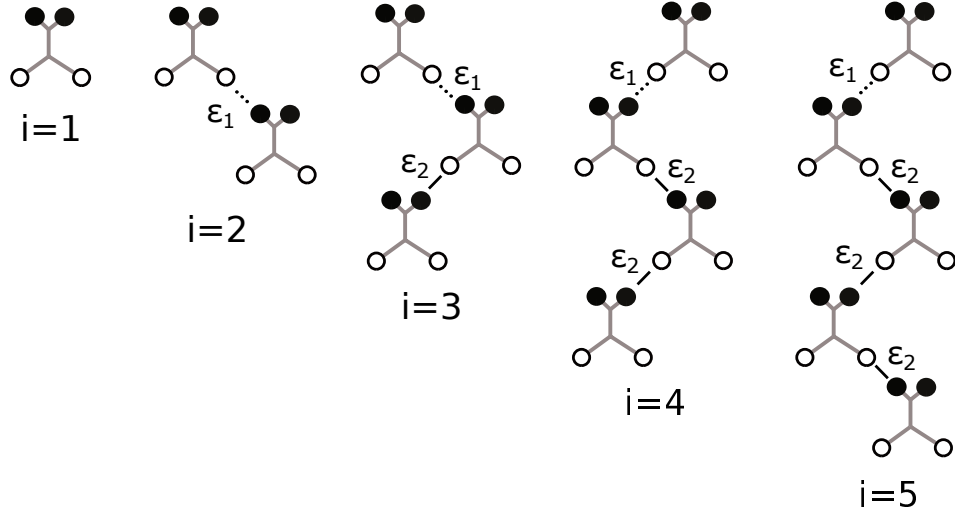


FIG. S49: Schematic representation of aggregates with size up to $i = 5$ in model 8. Only configurations corresponding to different energies are shown.

In this model it is straightforward to guess that

$$c_i = c_1^i 4^{i-1} K_1 K_2^{i-2}. \quad (\text{S126})$$

Then, the total concentration can be calculated as

$$c = \sum_{i=1}^{\infty} i c_i = c_1 + \frac{8c_1^2 K_1 (1 - 2c_1 K_2)}{(1 - 4c_1 K_2)^2}. \quad (\text{S127})$$

For the concentration of the free groups, $n_f = c - n_1 - n_2$, we have

$$n_f = c_1 + \frac{4K_1 c_1^2}{1 - 4K_2 c_1}. \quad (\text{S128})$$

1. Model 8m

Let us assume first that the 3530cm^{-1} peak corresponds to the out-of-phase vibrations of the NH_2 group in free molecules. In this case, we put $c_1 = Ax$ and substitute it into Equation S127.

The fitting results are shown in Figures S50 and S51. The quality of fit in Figure S50 can be characterized by $\text{AICc} = -501.5$, and the estimates of the model parameters yielded by the fits in Figure S51 are $\ln C_1 = -12.3 \pm 2 \ln[\text{l/mol}]$, $\epsilon_1 = -6.5 \pm 1 \text{ kcal/mol}$, $\ln C_2 = 0.1 \pm 0.6 \ln[\text{l/mol}]$, and $\epsilon_2 = 0.3 \pm 0.5 \text{ kcal/mol}$, with $r_1^2 = 0.89$ and $r_2^2 = 0.18$.

2. Model 8g

Next, it is assumed that the 3530cm^{-1} peak corresponds to absorption by free groups, so that $n_f = Ax$.

The fitting results are shown in Figures S52 and S53. The value of the AICc parameter characterizing the quality of fit in Figure S52 is -499.5 , and the estimates of the model parameters resulting from the fits in Figure S53 are $\ln C_1 = -13.6 \pm 3 \ln[\text{l/mol}]$, $\epsilon_1 = -7.3 \pm 1.6 \text{ kcal/mol}$, $\ln C_2 = -3.2 \pm 0.4 \ln[\text{l/mol}]$, and $\epsilon_2 = -2.0 \pm 0.2 \text{ kcal/mol}$, with $r_1^2 = 0.83$ and $r_2^2 = 0.95$.

3. Model 8s

In the case of the assumption $Ax = c_1 + 4K_1 c_1^2$ (NH_2 groups in unimers and dimers without ϵ_2 -bonds), the fitting results shown in Figures S54 and S55 are found. The quality of fit in Figure S54 can be characterized by

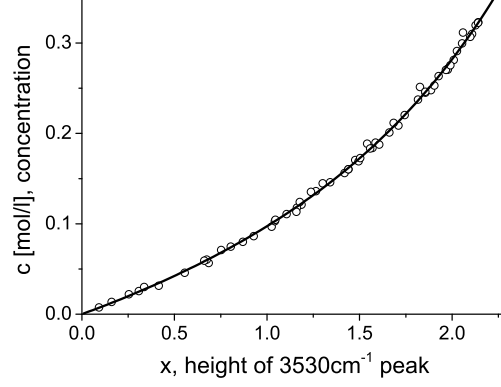


FIG. S50: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 8m.

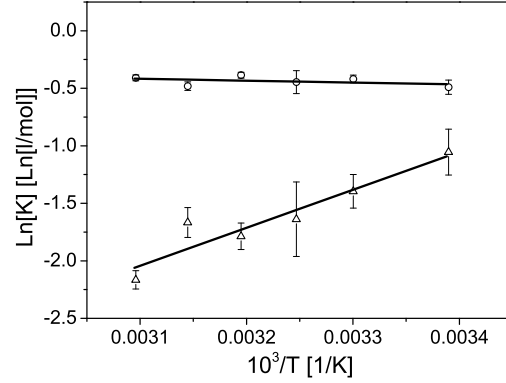


FIG. S51: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 8m.

$\text{AICc} = -501.4$. The estimates of the model parameters given by the fits in Figure S55 are $\ln C_1 = -14.2 \pm 2 \ln[\text{l/mol}]$, $\epsilon_1 = -7.9 \pm 1.4 \text{ kcal/mol}$, $\ln C_2 = -2.4 \pm 0.3 \ln[\text{l/mol}]$, and $\epsilon_2 = -1.3 \pm 0.2 \text{ kcal/mol}$, with $r_1^2 = 0.88$ and $r_2^2 = 0.93$.

J. Model 9

In this model, we allow two bonds per oxygen and only one bond per NH_2 group. We assume that the energy of the bond is determined by the bonding state of the oxygen. If the oxygen forms only one bond, then its energy is ϵ_1 , and if it forms two then its energy is ϵ_2 . This means that this model is analogous to model 5 with the roles of the acceptor and donor group exchanged. Therefore, all equations are the same, apart from the value of the number of free NH_2 groups. It is clear that in case of model 9 there is only one free group per each aggregate, so we have

$$n_f = \frac{1 - 4c_1K_1 - \sqrt{1 - 8c_1K_1 + 16c_1^2K_1^2 - 16c_1^2K_2^2}}{8c_1K_2^2}. \quad (\text{S129})$$

We do not repeat the fitting for models 9m and 9s as these cases are equivalent to models 5m and 5s.

1. Model 9g

We assume that the 3530cm^{-1} peak corresponds to absorption by free groups, so that $n_f = Ax$. The fitting results are not shown, because at higher temperatures they give a negative value of K_1 .

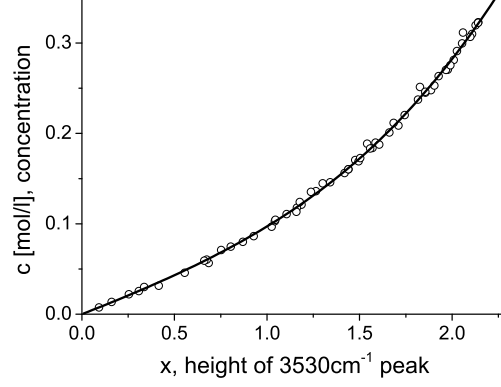


FIG. S52: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 8g.

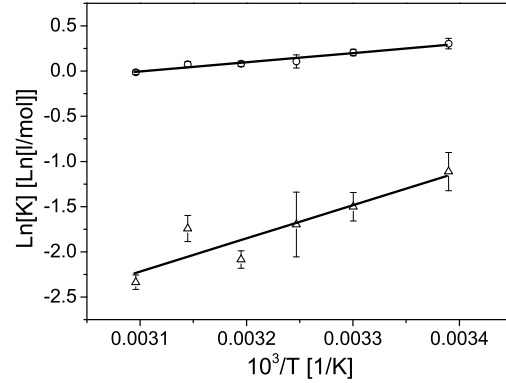


FIG. S53: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 8g.

K. Model 10

In this model, we allow two bonds per oxygen and only one bond per NH_2 group. We assume that the energy of the bond is determined by the bonding state of the NH_2 group in the acceptor molecule. If the hydrogen in the acceptor molecule is free, then the bond energy is ϵ_1 ; otherwise, it is ϵ_2 .

Therefore, this model is analogous to model 6 with the roles of the acceptor and donor groups exchanged, and all equations are the same, the only difference being the value of the number of free NH_2 groups. It is clear that, in the case of model 9, there is only one free group per each aggregate, so we have

$$n_f = c_1 + \frac{K_1 (1 - 4K_2c_1 + \sqrt{1 - 8c_1K_2})}{2K_2^2} + \frac{K_2^2 (1 - 8c_1K_2 + 8c_1^2K_2^2 - \sqrt{1 - 8c_1K_2} + 4c_1K_2\sqrt{1 - 8c_1K_2})}{8c_1K_2^4}. \quad (\text{S130})$$

1. Model 10g

Here, we assume that the 3530cm^{-1} peak corresponds to absorption by free groups, so that $n_f = Ax$.

The results of the fitting procedure are shown in Figures S56 and S57. The quality of fit in Figure S56 can be characterized by $\text{AICc} = -501.1$. The estimates of the model parameters for the fits in Figure S57 are $\ln C_1 = -11.3 \pm 2 \ln[\text{l/mol}]$, $\epsilon_1 = -6.3 \pm 1.0 \text{ kcal/mol}$, $\ln C_2 = -4.2 \pm 0.5 \ln[\text{l/mol}]$, and $\epsilon_2 = -2.4 \pm 0.3 \text{ kcal/mol}$, with

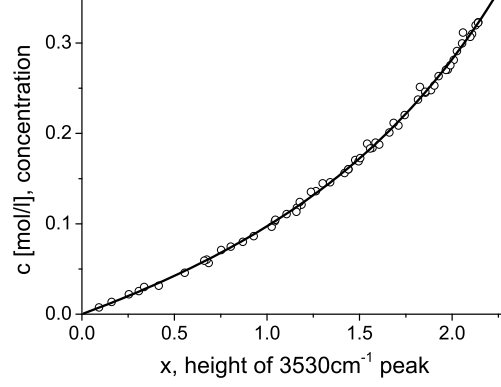


FIG. S54: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 8s.

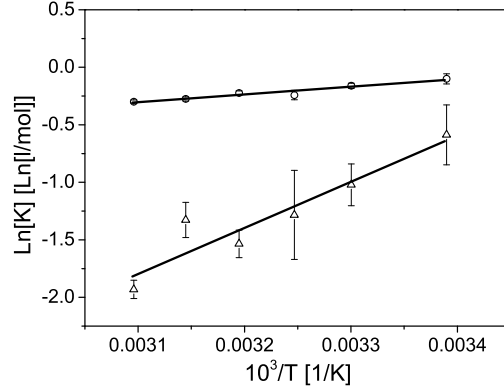


FIG. S55: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 8s.

$$r_1^2 = 0.90 \text{ and } r_2^2 = 0.95.$$

L. Model 11

In this model we allow two bonds per oxygen and only one bond per NH_2 group. We assume that the energy of the bond is determined by the bonding state of the acceptor in the donor molecules. If the oxygen in the donor molecule is free then the bond energy is ϵ_1 ; otherwise, it is ϵ_2 .

Therefore, this model is analogous to model 7 with the roles of the acceptor and donor group exchanged. All equations are the same, except for the value of the number of free NH_2 groups. Again, as in the two previous models, there is only one free group per aggregate

$$n_f = \frac{1 - 4K_2c_1 - 8K_2c_1^2(K_2 - K_1) - \sqrt{1 - 8K_2c_1 - 16K_2(K_1 - K_2)c_1^2}}{8K_2^2c_1}. \quad (\text{S131})$$

1. Model 11g

Here we assume that the 3530cm^{-1} peak corresponds to absorption by free groups, so that $n_f = Ax$. However, in this case, the fits do not converge.

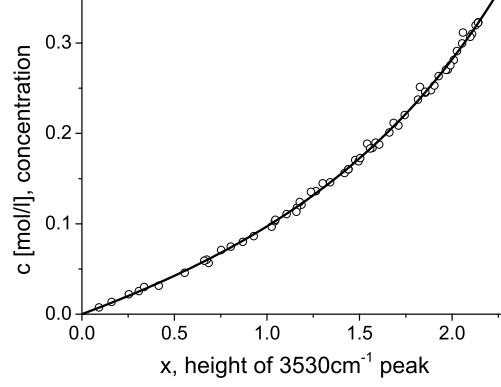


FIG. S56: Fit of the dependence of the total concentration on the height of the 3530cm^{-1} peak at $T = 22^\circ\text{C}$ with model 10g.

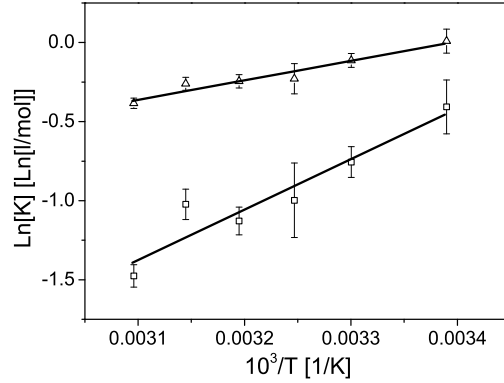


FIG. S57: Dependence of $\ln K_1$ (triangles) and $\ln K_2$ (circles) on $1/T$ for model 10g.

XII. THE BEST MODEL SELECTION

To compare the quality of fit of different models we use Akaike's information criterion. It states that the best model is that with the smallest value of

$$AIC = 2k - 2 \ln(L(A, K_1, K_2, \sigma)), \quad (\text{S132})$$

where $L(A, K_1, K_2, \sigma)$ is a likelihood function and k is the number of model parameters (4 in two-parameter models and 3 in one-parameter models because σ is also included in the context of the likelihood function). The likelihood function for the model with parameters obtained by minimization of the sum of squared deviations is usually written in the form

$$L(A, K_1, K_2, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - \text{model}(x_i, A, K_1, K_2))^2}{2\sigma^2} \right\}. \quad (\text{S133})$$

In this case, minimization of the sum of squared deviations is equivalent to maximization of the logarithm of the likelihood function. Then, for the AIC value we have

$$\begin{aligned} AIC &= 2k + n \ln(2\pi) + n \ln(\hat{\sigma}^2) + \frac{(y_i - \text{model}(x_i))^2}{2\hat{\sigma}^2} \\ &= 2k + n \ln(2\pi) + n \ln \frac{\text{RSS}}{n} + n, \end{aligned} \quad (\text{S134})$$

where the estimate for $\hat{\sigma}$ is $\hat{\sigma}^2 = \text{RSS}/n$ and the sum of the squares of the residuals is $\text{RSS} = \sum_{i=1}^n (y_i - \text{model}(x_i))^2$.

For a finite sample size there exists the following correction:

$$\text{AICc} = \text{AIC} + \frac{2(k+1)(k+2)}{n-k-2}, \quad (\text{S135})$$

which is useful when we compare models with different number of parameters.

The relative probabilities of the two models with values of the information criterion given by AIC_1 and AIC_2 respectively can be estimated as

$$\exp\left(\frac{\text{AIC}_1 - \text{AIC}_2}{2}\right) \quad (\text{S136})$$

So, if $\text{AIC}_1 - \text{AIC}_2 = -501.7 + 493.9 = -7.8$ then model with value AIC_1 is 49.4 times more probable than the model with value AIC_2 .